

STIC-Biotech/ChemLib

88752

From: Davis, Minh-Tam
Sent: Tuesday, March 11, 2003 3:01 PM
To: STIC-Biotech/ChemLib
Subject: Search request for 09/847046

Please search in commercial database, PGPUB, and issued patent files:

1) SEQ ID NO:1 and 2.

Thank you.

MINH TAM DAVIS

ART UNIT 1642, ROOM 8A01, MB 8E12

305-2008

Oct 2000
priority date

Edward Hart
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

RECEIVED
MAR 11 2003
STIC/Biotech
(STIC)

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 3/13/03
Date Completed: 3/25/03
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: 8/10/04
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (US)

PA (E0SB-) EOS BIOTECHNOLOGY INC.
XX
PI Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;
XX
DR WPI: 2002-471335/50.
DR P-PSDB: AB661812.
XX
PT Detecting a prostate cancer-associated transcript in a cell in a
modulator of PC, by determining if prostate cancer-associated genes
are expressed in a prostate tissue
XX
PS Claim 22: Page 310-311: 436pp; English.
XX
CC The present invention relates to methods of detecting a prostate
cancer-associated transcript in a cell from a patient. The method
comprises contacting a biological sample from the patient with
CC prostate cancer-associated polynucleotides (designated PC genes) that
selectively hybridise to a sequence that is at least 80% identical
CC to them. The prostate cancer-associated polynucleotide sequences
are differentially expressed in prostate tumour tissue or in
CC prostate cancer and are derived from the tissues of various
CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).
CC The methods of the invention are useful for diagnosing and treating
CC prostate cancer in mammals. The prostate cancer-associated genes are
CC useful for diagnosing or treating prostate cancer, as well as for
CC identifying modulators of prostate cancer or agents that inhibit
CC prostate cancer. The nucleic acid sequences are particularly useful
CC in gene therapy, as a vaccine or in antisense applications.
CC ABK92115-ABK92263 represent prostate cancer-associated polynucleotide
sequences.
XX
SQ Sequence 4526 BP; 1444 A; 817 C; 890 G; 1374 T; 1 other:
Query Match 100.0%; Score 4525.6; DB 24; Length 4526;
Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Gaps 0;
Matches 4526; Conservative 0; Indels 0; Gaps 0;
QY 1 GCCGAGTCGGTGGCGGCTGACAGGCTGGAGGAGAGTGTACGCCCTTTCAGAGTTGGCG 60
DB 1 GCCGAGTCGGTGGCGGCTGACAGGCTGGAGGAGAGTGTACGCCCTTTCAGAGTTGGCG 60
QY 61 AAGTGGTCCAGGCTACCCGGCTAGTGTGGCAGGCGCCCTTCTCTCTCTCTCTCTCT 120
DB 61 AAGTGGTCCAGGCTACCCGGCTAGTGTGGCAGGCGCCCTTCTCTCTCTCTCTCTCT 120
QY 121 CGGCTGGCGGCGGAGTGTGGCGGCGGCGCTCGGGAACGGCCAGGTCGCCGCCG 180
DB 121 CGGCTGGCGGCGGAGTGTGGCGGCGGCGCTCGGGAACGGCCAGGTCGCCGCCG 180
QY 181 AGGTCCCGGCGAGTAACATAGATCATAGTAAACCTTCTGAAGTTGTTCAGAAAA 240
DB 181 AGGTCCCGGCGAGTAACATAGATCATAGTAAACCTTCTGAAGTTGTTCAGAAAA 240
QY 241 ATTTGAAAGTAAATTAATAAATAAGATTAAACGACATACAGAGACACACATGA 300
DB 241 ATTTGAAAGTAAATTAATAAATAAGATTAAACGACATACAGAGACACACATGA 300
QY 301 AGTGTGTCTTAGGAACACAGACAGACAGTGAAGAAACAGACAAATCCGCTCAGATAC 360
DB 301 AGTGTGTCTTAGGAACACAGACAGACAGTGAAGAAACAGACAAATCCGCTCAGATAC 360
QY 361 AACTGACGATTAATGTTTTCGGGCTCAATGTCTTTAGAGTTGGAGTCTCTTTTGTCA 420
DB 361 AACTGACGATTAATGTTTTCGGGCTCAATGTCTTTAGAGTTGGAGTCTCTTTTGTCA 420
QY 421 TAATGTGCATTTTACATGCGCAACATTAACCTTTACAGAACTGAGTCCACAGAAAT 480
DB 421 TAATGTGCATTTTACATGCGCAACATTAACCTTTACAGAACTGAGTCCACAGAAAT 480
QY 481 ATTTTACATTCGACACAGGCTTTGAAGAACTGAGTGGGCTTGAAGCCTTGAGG 540
DB 481 ATTTTACATTCGACACAGGCTTTGAAGAACTGAGTGGGCTTGAAGCCTTGAGG 540

QY 541 ACTATGGAATTTGAGTTGCCAAGTTAAATGTGTCAAGAAGAAATATCAAGATACTGTG 600
DB 541 ACTATGGAATTTGAGTTGCCAAGTTAAATGTGTCAAGAAGAAATATCAAGATACTGTG 600
QY 601 GAAAAGAAAAGATTTGATGAAGCATATTTATTCAGAGGCAACATTTGCTCAGAGAA 660
DB 601 GAAAAGAAAAGATTTGATGAAGCATATTTATTCAGAGGCAACATTTGCTCAGAGAA 660
QY 661 TCCCTACTGACCTTGTGATGTGAATGCCAATTCGCGCAATGTCCTTGGCTCTTC 720
DB 661 TCCCTACTGACCTTGTGATGTGAATGCCAATTCGCGCAATGTCCTTGGCTCTTC 720
QY 721 TTTTATGTAAGTGAATATATTTATTCACCACTGGAAGACCTTCAGAACATAGAAATGCTC 780
DB 721 TTTTATGTAAGTGAATATATTTATTCACCACTGGAAGACCTTCAGAACATAGAAATGCTC 780
QY 781 TGAAGAGAAAAGCAATATTTATTTCTCATATGTAGAGCCATTTGAAATCCAGAGCACA 840
DB 781 TGAAGAGAAAAGCAATATTTATTTCTCATATGTAGAGCCATTTGAAATCCAGAGCACA 840
QY 841 GAGAGTCATGGAAGCCGTTTGTGTATGAGGACTACATCAATTTGCTTAACACAG 900
DB 841 GAGAGTCATGGAAGCCGTTTGTGTATGAGGACTACATCAATTTGCTTAACACAG 900
QY 901 AAATGGCCCTTTTGAAGATTTGGCTGTAGAGATGTGAATATGCACATCTACTTTT 960
DB 901 AAATGGCCCTTTTGAAGATTTGGCTGTAGAGATGTGAATATGCACATCTACTTTT 960
QY 961 TTTATTTAACTAGTCTTGAGCTTGACCCAGCAATGTATGAACACATTAAGAACAGC 1020
DB 961 TTTATTTAACTAGTCTTGAGCTTGACCCAGCAATGTATGAACACATTAAGAACAGC 1020
QY 961 TTTATTTAACTAGTCTTGAGCTTGACCCAGCAATGTATGAACACATTAAGAACAGC 1020
DB 961 TTTATTTAACTAGTCTTGAGCTTGACCCAGCAATGTATGAACACATTAAGAACAGC 1020
QY 1021 CATTTGACTACATGAACTTACCTGTTTATTAAGAAATGAAGACCTCTTGTGACTG 1080
DB 1021 CATTTGACTACATGAACTTACCTGTTTATTAAGAAATGAAGACCTCTTGTGACTG 1080
QY 1081 AAGTGTGTAAGATCTCTCAACAGTTTCAACCTGCTTCAACCTGGGCTTACCACTGG 1140
DB 1081 AAGTGTGTAAGATCTCTCAACAGTTTCAACCTGCTTCAACCTGGGCTTACCACTGG 1140
QY 1141 TTTTATTTATTTAGCAACAGGCTACTATGAACTGTATGAACACATGAAATGGTGG 1200
DB 1141 TTTTATTTATTTAGCAACAGGCTACTATGAACTGTATGAACACATGAAATGGTGG 1200
QY 1201 CTGGCGCTTCTTGGGAAAAGCAGAGATCTTACTCTTGTAAAGGACTCTTGTGAAGTGA 1260
DB 1201 CTGGCGCTTCTTGGGAAAAGCAGAGATCTTACTCTTGTAAAGGACTCTTGTGAAGTGA 1260
QY 1261 ACATTCCTCAAGATGCTAATGTGTCTTCAAAAAGCAGAGAGGAGTTCAGATGAAT 1320
DB 1261 ACATTCCTCAAGATGCTAATGTGTCTTCAAAAAGCAGAGAGGAGTTCAGATGAAT 1320
QY 1321 TTTTGGTATTAACATGATGATTTTAAATATCTCATGAGGAAAATTAATGACACTGG 1380
DB 1321 TTTTGGTATTAACATGATGATTTTAAATATCTCATGAGGAAAATTAATGACACTGG 1380
QY 1381 AGGAATATCAAGAAAGATGAAGACATGACATGGAAGGTCAGATATAGATGTTCCAGATG 1440
DB 1381 AGGAATATCAAGAAAGATGAAGACATGACATGGAAGGTCAGATATAGATGTTCCAGATG 1440
QY 1441 ATGAAGTGGCAGAACTGTTTTCAGAGATAGGAAGAGAAATTAATCTTTGAACTTACAG 1500
DB 1441 ATGAAGTGGCAGAACTGTTTTCAGAGATAGGAAGAGAAATTAATCTTTGAACTTACAG 1500
QY 1501 TGAAGCTTAAGAAAGAACTTTAATGCAACAGATGAGTCTGACAGATAGTACTCT 1560
DB 1501 TGAAGCTTAAGAAAGAACTTTAATGCAACAGATGAGTCTGACAGATAGTACTCT 1560
QY 1561 TCTATGCTGTGGCAAGCAGTATCCATGSCATTTTTCGAATCTATATGATGTGGCAG 1620
DB 1561 TCTATGCTGTGGCAAGCAGTATCCATGSCATTTTTCGAATCTATATGATGTGGCAG 1620

| | | | |
|----|------|--|------|
| QY | 1621 | TTAACTGAAAGGACACTTCTACTATGCTTCTTACTAGATAAAGCTGTGACATGTGCTG | 1680 |
| Db | 1621 | TTAAACTGAAAGGACACTTCTACTATGCTTCTTACTAGATAAAGCTGTGACATGTGCTG | 1680 |
| QY | 1681 | ATGTATGTACTAAGCAAAAGTTTACTAGAAATTTCTTCACTTAAAGATGTACAAGAAAGCG | 1740 |
| Db | 1681 | ATGTATGTACTAAGCAAAAGTTTACTAGAAATTTCTTCACTTAAAGATGTACAAGAAAGCG | 1740 |
| QY | 1741 | AGAACCCAGTATCTTATGCTGGATGTTAGGAACCAAGATCTCTTAAATTTTATCCAGC | 1800 |
| Db | 1741 | AGAACCCAGTATCTTATGCTGGATGTTAGGAACCAAGATCTCTTAAATTTTATCCAGC | 1800 |
| QY | 1801 | TCAACAGGATTTCTATTCACAGTGAATATACATCCACAAGACAGAGATATTTAA | 1860 |
| Db | 1801 | TCAACAGGATTTCTATTCACAGTGAATATACATCCACAAGACAGAGATATTTAA | 1860 |
| QY | 1861 | GTGGGGAATTTATTAAGACCTCATCTGTATTTCTATGCTGTGACATTTGGACATATTA | 1920 |
| Db | 1861 | GTGGGGAATTTATTAAGACCTCATCTGTATTTCTATGCTGTGACATATTTGGACATATTA | 1920 |
| QY | 1921 | GTCCAACCATGAAAAACAGCAAAAGAAATTTAGTAACAGAGAAACTACTTAAAGAT | 1980 |
| Db | 1921 | GTCCAACCATGAAAAACAGCAAAAGAAATTTAGTAACAGAGAAACTACTTAAAGAT | 1980 |
| QY | 1981 | ATGTATATCATCGGAATTTATTTGGAAGAGATTTTGTCTACTGTCAACCAATATGCTG | 2040 |
| Db | 1981 | ATGTATATCATCGGAATTTATTTGGAAGAGATTTTGTCTACTGTCAACCAATATGCTG | 2040 |
| QY | 2041 | CAAGCTTCCAGCCCTCTCTTGGCAGACACACAGAAAGCAAAATAGAGACATCCAC | 2100 |
| Db | 2041 | CAAGCTTCCAGCCCTCTCTTGGCAGACACACAGAAAGCAAAATAGAGACATCCAC | 2100 |
| QY | 2101 | TAGTAGCACACATGCAACAAGACATAGTTCAATATAAACAAGATCAGTACTGGAAATGT | 2160 |
| Db | 2101 | TAGTAGCACACATGCAACAAGACATAGTTCAATATAAACAAGATCAGTACTGGAAATGT | 2160 |
| QY | 2161 | TTCCGGAAATCAGTGTGAAAAATCTTCCAGTATTTTACAGACTTCAGAAACCATATTTGA | 2220 |
| Db | 2161 | TTCCGGAAATCAGTGTGAAAAATCTTCCAGTATTTTACAGACTTCAGAAACCATATTTGA | 2220 |
| QY | 2221 | TTTTTGTACAGATGGGACGTAAATCCCAATATTAATAAACAATTTGGACATCGTAA | 2280 |
| Db | 2221 | TTTTTGTACAGATGGGACGTAAATCCCAATATTAATAAACAATTTGGACATCGTAA | 2280 |
| QY | 2281 | AGCAGAATATCTTGGATTCATTACTCCCAATGCTGTAAATTAAGAATATCTCCAGTGG | 2340 |
| Db | 2281 | AGCAGAATATCTTGGATTCATTACTCCCAATGCTGTGTAAATTAAGAATATCTCCAGTGG | 2340 |
| QY | 2341 | GGAGAGGAATCTTGGCGGCAATATTTTGATCTCTGTGCTCCCTCTCTTGTGTTGG | 2400 |
| Db | 2341 | GGAGAGGAATCTTGGCGGCAATATTTTGATCTCTGTGCTCCCTCTCTTGTGTTGG | 2400 |
| QY | 2401 | TGAATCTGCATTCAGGTGGCCAAAGTATTTGCATTTCTTCAGACACAGGCTATTAATGAG | 2460 |
| Db | 2401 | TGAATCTGCATTCAGGTGGCCAAAGTATTTGCATTTCTTCAGACACAGGCTATTAATGAG | 2460 |
| QY | 2461 | AAAACCTTGATTTGTGCTCAGAAATTTGAAGACAGAGATAGAAAATCATATTCACAATTT | 2520 |
| Db | 2461 | AAAACCTTGATTTGTGCTCAGAAATTTGAAGACAGAGATAGAAAATCATATTCACAATTT | 2520 |
| QY | 2521 | TACCTGCTCAAGATGGAAACCTCTCTCCAGCTTATGATTTTCTTAAGTATGATAGATG | 2580 |
| Db | 2521 | TACCTGCTCAAGATGGAAACCTCTCTCCAGCTTATGATTTTCTTAAGTATGATAGATG | 2580 |
| QY | 2581 | CCGCAACATCTCAACGTGGCAGCTAGGAAGATTTCCCAAGTGTGAAAGAAACACATGTGC | 2640 |
| Db | 2581 | CCGCAACATCTCAACGTGGCAGCTAGGAAGATTTCCCAAGTGTGAAAGAAACACATGTGC | 2640 |
| QY | 2641 | AGGAGAAATGATTAAGGAACACATGAAGATTAATGGCAGCTCGAAGAAAGAACCCGATTGAA | 2700 |
| Db | 2641 | AGGAGAAATGATTAAGGAACACATGAAGATTAATGGCAGCTCGAAGAAAGAACCCGATTGAA | 2700 |
| QY | 2701 | CTCTGACAAATAAAGCATTTGGAAATAGAAATTTGGTTTAAAGAACAGAAAAATCATATTTA | 2760 |

| | | | |
|----|------|---|------|
| Dd | 2701 | CTCTGAGATTAACCATTTGGAAATAAGATATTGGTTTTTAAAGACAGCAAAAATCATTTTAA | 2760 |
| Oy | 2761 | GACGTGAATAAAGACTTGATAGATGCCTCAAABSTGAACATAATTTTATPAGGGCTGGTGTTC | 2820 |
| Dd | 2761 | GACGTGAATAAAGACTTGATAGATGCCTCAAABSTGAACATAATTTTATPAGGGCTGGTGTTC | 2820 |
| Oy | 2821 | AAATTTTGGCATGTATAGACTTAATTTATTTTCCTTAAAGAATATATTAATCATTTTC | 2880 |
| Dd | 2821 | AAATTTTGGCATGTATAGACTTAATTTATTTTCCTTAAAGAATATATTAATCATTTTC | 2880 |
| Oy | 2881 | AAGTTTGACAGCTAGTGGCATTCCAAATBGAATTTATATAACTCACATRTTTATTTHAA | 2940 |
| Dd | 2881 | AAGTTTGACAGCTAGTGGCATTCCAAATBGAATTTTAAATAACTCACATRTTTATTTHAA | 2940 |
| Oy | 2941 | ATTTCCTGTATACACATTTTAAACAAGTAAASTGAGCGGCCAAATATTTTGATATTT | 3000 |
| Dd | 2941 | ATTTCCTGTATACACATTTTAAACAAGTAAASTGAGCGGCCAAATATTTTGATATTT | 3000 |
| Oy | 3001 | ACTTTTCACCAGTAGATATACCCAAAAATGCGAAATATAGAAATTTATTAATGATATTT | 3060 |
| Dd | 3001 | ACTTTTCACCAGTAGATATACCCAAAAATGCGAAATATAGAAATTTATTAATGATATTT | 3060 |
| Oy | 3061 | TACATCTTTTTTGTACCAAGTCTTCTAATATGACATATTTTATATCTTACATCTGATTT | 3120 |
| Dd | 3061 | TACATCTTTTTTGTACCAAGTCTTCTAATATGACATATTTTATATCTTACATCTGATTT | 3120 |
| Oy | 3121 | CTTACCTCCGAGTAGGCATATTTTCAAGTGTCTCAATGCCACATGTGGCCGTGACTACGT | 3180 |
| Dd | 3121 | CTTACCTCCGAGTAGGCATATTTTCAAGTGTCTCAATGCCACATGTGGCCGTGACTACGT | 3180 |
| Oy | 3181 | ATTGGACAGTTCAGTACTAGACAAAACTAGCATTAATTAATTAATTAATTAATTAATTAAT | 3240 |
| Dd | 3181 | ATTGGACAGTTCAGTACTAGACAAAACTAGCATTAATTAATTAATTAATTAATTAATTAAT | 3240 |
| Oy | 3241 | CTATTTGGATTAAATTAACCTAATCACAGTTAACTCCACAGTGCATTTATGACGTG | 3300 |
| Dd | 3241 | CTATTTGGATTAAATTAACCTAATCACAGTTAACTCCACAGTGCATTTATGACGTG | 3300 |
| Oy | 3301 | ACAGTTATATTTGTTTTATTGGAATCATGATTAATTAATTAATTAATTAATTAATTAATTAAT | 3360 |
| Dd | 3301 | ACAGTTATATTTGTTTTATTGGAATCATGATTAATTAATTAATTAATTAATTAATTAATTAAT | 3360 |
| Oy | 3361 | GATATTTAGCAATTTGTGGGAGACATTTTGATGTCATACATAGGCGCATTAATTGACATT | 3420 |
| Dd | 3361 | GATATTTAGCAATTTGTGGGAGACATTTTGATGTCATACATAGGCGCATTAATTGACATT | 3420 |
| Oy | 3421 | TAGTGAGTAGAGGCCATGATCCTGCTAAATACTGCAATGAGACAGCGCCCCACACAA | 3480 |
| Dd | 3421 | TAGTGAGTAGAGGCCATGATCCTGCTAAATACTGCAATGAGACAGCGCCCCACACAA | 3480 |
| Oy | 3481 | AGAATTAATCCCGCCGAAATGATAGTGTGCTGCAAGGCTAGTAACCTTGTTGTTAAAGTA | 3540 |
| Dd | 3481 | AGAATTAATCCCGCCGAAATGATAGTGTGCTGCAAGGCTAGTAACCTTGTTGTTAAAGTA | 3540 |
| Oy | 3541 | ACCTGTGGCAGACTAGAGTTTCCAGAAATTTCTGCTGCTACGATCATGTTTGAAAA | 3600 |
| Dd | 3541 | ACCTGTGGCAGACTAGAGTTTCCAGAAATTTCTGCTGCTACGATCATGTTTGAAAA | 3600 |
| Oy | 3601 | AAATTTGGCATTAATAAGATATGATTAAGATGCTTATCCTGATTAATTAATTAATTAATTAAT | 3660 |
| Dd | 3601 | AAATTTGGCATTAATAAGATATGATTAAGATGCTTATCCTGATTAATTAATTAATTAATTAAT | 3660 |
| Oy | 3661 | CTTGATCTTTTCTATATTTTTCAGAAAGTATGGGATTAACCTCAAAAGAGACTACAGAT | 3720 |
| Dd | 3661 | CTTGATCTTTTCTATATTTTTCAGAAAGTATGGGATTAACCTCAAAAGAGACTACAGAT | 3720 |
| Oy | 3721 | GATATTTATATTTTAATGATGATCTTAAACCTGCTTATTTCTCAAGATTAATGAGCTA | 3780 |
| Dd | 3721 | GATATTTATATTTTAATGATGATCTTAAACCTGCTTATTTCTCAAGATTAATGAGCTA | 3780 |
| Oy | 3781 | AAATTCAGATTTGAACAGGATTCAGCATTCCTGCATCTCTCATGGAAGAGAGCTGCC | 3840 |

```
Db 3781 AATTTGAGATTGACAGGATTCAGACATCTGCCATCTCCTATGAAAGAGCGTCCC 3840
Qy 3841 TCATCTGAAGCGTCTCTGAATCTACCCCTGCAAGCTTCAGACAAATCAGTGTCTCCC 3900
Db 3841 TCATCTGAAGCGTCTCTGAATCTACCCCTGCAAGCTTCAGACAAATCAGTGTCTCCC 3900
Qy 3901 TGAGCCACACGCGCTCTCTGTGAGGAGGAGAAAGATTAGCCAAAGAGTTAATTTTCAT 3960
Db 3901 TGAGCCACACGCGCTCTCTGTGAGGAGGAGAAAGATTAGCCAAAGAGTTAATTTTCAT 3960
Qy 3961 TCCAAATCATTAGCTGTTAGAGTATCTGTTGTAGCAGTGTGTTGTCTCATTTTTCAT 4020
Db 3961 TCCAAATCATTAGCTGTTAGAGTATCTGTTGTAGCAGTGTGTTGTCTCATTTTTCAT 4020
Qy 4021 CTGTGCAATTTTGTGAGCATTGTGTGAAATATTTCTATTTGGTCTCTACTGTAATTTTC 4080
Db 4021 CTGTGCAATTTTGTGAGCATTGTGTGAAATATTTCTATTTGGTCTCTACTGTAATTTTC 4080
Qy 4081 TTTTAAATCTACTGATATCTTGTCTTTAAATTTTCTTCACATATGCTTGCCTGAT 4140
Db 4081 TTTTAAATCTACTGATATCTTGTCTTTAAATTTTCTTCACATATGCTTGCCTGAT 4140
Qy 4141 ACAAGCATTTTATACGTAATTAAGAAATCTAACAGCTAACAGCTAACAGTATGTCAT 4200
Db 4141 ACAAGCATTTTATACGTAATTAAGAAATCTAACAGCTAACAGCTAACAGTATGTCAT 4200
Qy 4201 MTATTTCTTATACATAGACCCGTTGCTACTCTCAGCACCCCTCTCCTCAATTTTTC 4260
Db 4201 MTATTTCTTATACATAGACCCGTTGCTACTCTCAGCACCCCTCTCCTCAATTTTTC 4260
Qy 4261 CTGTGAGATGATGCTGATTAATACATCTTTCATTTTCTTTTATTTAATATGAGAA 4320
Db 4261 CTGTGAGATGATGCTGATTAATACATCTTTCATTTTCTTTTATTTAATATGAGAA 4320
Qy 4321 CAATGAGAGGAGACTCTAATATATAGTGTAGTAATTAACATCATATAGCTAATTTTA 4380
Db 4321 CAATGAGAGGAGACTCTAATATATAGTGTAGTAATTAACATCATATAGCTAATTTTA 4380
Qy 4381 GAAATGCTAATTAAGTACGACGACATAGAACATGAATGCTTACTGATTTGACCTTT 4440
Db 4381 GAAATGCTAATTAAGTACGACGACATAGAACATGAATGCTTACTGATTTGACCTTT 4440
Qy 4441 GTGAGCAATTTTGACAGTCAATTTGTTGTCATTAATTAATTAATGCTGCGGTTTC 4500
Db 4441 GTGAGCAATTTTGACAGTCAATTTGTTGTCATTAATTAATTAATGCTGCGGTTTC 4500
Qy 4501 AGAATACCTTCAAAAAA 4526
Db 4501 AGAATACCTTCAAAAAA 4526

RESULT 2
AAH15799
ID AAH15799 standard: cDNA: 2820 BP.
XX
AC AAH15799;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:14261.
XX
KW Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
```

```
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
PS Claim 8: SEQ ID 14261; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises at least 15 nucleotides and the combination
CC of the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB9246 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2820 BP; 911 A; 496 C; 504 G; 909 T; 0 other:
Query Match 62.1%; Score 2811.6; DB 22; Length 2820;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2814; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
Qy 1686 TGTACTAGCAAAATGTTACTGTAATTTCTATCATTAAGATGACAAAGGCGGAAAC 1745
Db 1 TGTACTAGCAAAATGTTACTGTAATTTCTATCATTAAGATGACAAAGGCGGAAAC 60
Qy 1746 CCAATATCTTATGCTGGAATGTTAGAACCAAGATCTCTAAATTTATCCAGCTCAAC 1805
Db 61 CCAATATCTTATGCTGGAATGTTAGAACCAAGATCTCTAAATTTATCCAGCTCAAC 120
Qy 1806 AGGATTTCAATATCAGTGAATATACATGATGCCAAGACAGCAAGATATTTAATGAG 1865
Db 121 AGGATTTCAATATCAGTGAATATACATGATGCCAAGACAGCAAGATATTTAATGAG 180
Qy 1866 GAATTTATTAAGACCCATCTTGTATTCAGTGTGACATTTGGAGCATTTAGTCCA 1925
Db 181 GAATTTATTAAGACCCATCTTGTATTCAGTGTGACATTTGGAGCATTTAGTCCA 240
Qy 1926 ACCATGAAAACAGCAAAAGAGATTTTGTAGAGAGGAAAGTCTTAAAGAGATATGTT 1985
Db 241 ACCATGAAAACAGCAAAAGAGATTTTGTAGAGAGGAAAGTCTTAAAGAGATATGTT 300
Qy 1986 ATCACTGGAATTTATCTGAGAAGATGTTTGTCTACTGTCAACCAATATGCTGCAAGT 2045
Db 301 ATCACTGGAATTTATCTGAGAAGATGTTTGTCTACTGTCAACCAATATGCTGCAAGT 360
Qy 2046 CTTCAGGCCCTGCTGCTGCGACAGACACAGCAAGCAAAATAGAGAGCATCCACTAGCT 2105
Db CTTCAGGCCCTGCTGCTGCGACAGACACAGCAAGCAAAATAGAGAGCATCCACTAGCT 2105
```

| | | | | |
|---|---|------|--|------|
| D | b | 361 | CTTCCAGCCTGCTGCTTGGCCACACACACAGAGCAAAATAGAGCATCCCACTAGCT | 420 |
| Q | y | 2106 | AGCACATGACACAAAGACATAGTTCAAATTAATACGATGACCTACTGGAATGTTCCG | 2165 |
| D | b | 421 | AGCACACATGACAAAGACATAGTTCAAATTAATACGATGACCTACTGGAATGTTCCG | 480 |
| Q | y | 2166 | GAATACACTGTGGAAAATCTTCCAGCTATTTTCAGACTTCAGAAAACATATTGATTTTG | 2225 |
| D | b | 481 | GAATACACTGTGGAAAATCTTCCAGCTATTTTCAGACTTCAGAAAACATATTGATTTTG | 540 |
| Q | y | 2226 | TTTCGATAGGCACTGTAAATTCCTCAATTTAAAAAAGCAATTTTCACACTGTAAAGAG | 2285 |
| D | b | 541 | TTTCGATAGGCACTGTAAATTCCTCAATTTAAAAAAGCAATTTTCACACTGTAAAGAG | 600 |
| Q | y | 2286 | AAATACCTGGATTCATTTACTTCATCTGTTTAAATCTTAAACAATACACAGTGGGAGA | 2345 |
| D | b | 601 | AAATACCTGGATTCATTTACTTCATCTGTTTAAATCTTAAACAATACCTGAGTGGGAGA | 660 |
| Q | y | 2346 | GGAACTCTGGGGGCATATTTTGATTCCTTGCCCTCCCTCTCTTCTGTTTGGTGAAT | 2405 |
| D | b | 661 | GGAACTCTGGGGGCATATTTTGATTCCTTGCCCTCCCTCTCTTCTGTTTGGTGAAT | 720 |
| Q | y | 2406 | CTGATTCAGTGGGCCAAGATTTTGATTTCCCTCAGACACAGGCTAATTTGAGAAAAC | 2465 |
| D | b | 721 | CTGATTCAGTGGGCCAAGATTTTGATTTCCCTCAGACACAGGCTAATTTGAGAAAAC | 780 |
| Q | y | 2466 | CTTGTATTGTGGCGTGAAGAAATTAGAAGCAGACTAGAAAATCATATCACATTTTACCT | 2525 |
| D | b | 781 | CTTGTATTGTGGCGTGAAGAAATTAGAAGCAGACTAGAAAATCATATCACATTTTACCT | 840 |
| Q | y | 2526 | GCTCAGAAATGGAACCTCCTCTTCCAGCTTATGATTTTCTAAGTATGATAGTCCGCA | 2585 |
| D | b | 841 | GCTCAGAAATGGAACCTCCTCTTCCAGCTTATGATTTTCTAAGTATGATAGTCCGCA | 900 |
| Q | y | 2586 | ACATCTCAACGTGGCACTAGSAAAGTTCCCAAGTATATAAAGAAACAGATGTGCAGAG | 2645 |
| D | b | 901 | ACATCTCAACGTGGCACTAGSAAAGTTCCCAAGTATATAAAGAAACAGATGTGCAGAG | 960 |
| Q | y | 2646 | AATGATTAAGGAACAACATGGAATTAATGGCAGTCAGAAAAGAACGATTTGAAACTGTG | 2705 |
| D | b | 961 | AATGATTAAGGAACAACATGGAATTAATGGCAGTCAGAAAAGAACGATTTGAAACTGTG | 1020 |
| Q | y | 2706 | AGAAATTAAGCATTGGAATAGAACTAATTTGTTTAAAGAACGAAAATCATTTAGACGT | 2765 |
| D | b | 1021 | AGAAATTAAGCATTGGAATAGAACTAATTTGTTTAAAGAACGAAAATCATTTAGACGT | 1080 |
| Q | y | 2766 | GATTAAGAATTTAGAGTGTCTAAAAAGCAACTAATTTTATAGGGCTGTGTTCCAAATTT | 2825 |
| D | b | 1081 | GATTAAGAATTTAGAGTGTCTAAAAAGCAACTAATTTTATAGGGCTGTGTTCCAAATTT | 1140 |
| Q | y | 2826 | TTTTTGGCATGATGAGACTTAATTTATTTTCCCTTAAGAAATTAATTAATCAATTCAGATT | 2885 |
| D | b | 1141 | TTTTTGGCATGATGAGACTTAATTTATTTTCCCTTAAGAAATTAATTAATCAATTCAGATT | 1200 |
| Q | y | 2886 | TGCAGACTAGTGCATCACTCAATAGAAATTAATTAATTAAGTCCATATTTTATTTAAATTTT | 2945 |
| D | b | 1201 | TGCAGACTAGTGCATCACTCAATAGAAATTAATTAATTAAGTCCATATTTTATTTAAATTTT | 1260 |
| Q | y | 2946 | CTATATACCTACATTTAAACAAGTAAGTAAGGACGAGGCAAAATATTTTGATATTACTTT | 3005 |
| D | b | 1261 | CTATATACCTACATTTAAACAAGTAAGTAAGGACGAGGCAAAATATTTTGATATTACTTT | 1320 |
| Q | y | 3006 | TCACCCAGTGTATATCCCAAAATAGGGAATTAATTAAGAAATTAATTAAGATATTTTACAT | 3065 |
| D | b | 1321 | TCACCCAGTGTATATCCCAAAATAGGGAATTAATTAAGAAATTAATTAAGATATTTTACAT | 1380 |
| Q | y | 3066 | CTTTTTTTGTACCAAGTCTTCTAAATGACGATACATATTTTATATCTACGATTTCTTAC | 3125 |
| D | b | 1381 | CTTTTTTTGTACCAAGTCTTCTAAATGACGATACATATTTTATATCTACGATTTCTTAC | 1440 |
| Q | y | 3126 | TTCCGAGTAGGCATATTTTCAAGTGTTCATTTGCCACATGTGGCTGTGACTACTGTATTTGG | 3185 |
| D | b | 1441 | TTCCGAGTAGGCATATTTTCAAGTGTTCATTTGCCACATGTGGCTGTGACTACTGTATTTGG | 1500 |

| | | | |
|----|------|--|------|
| QY | 3186 | ACAGTTCAGTACTAGACAAAACAGCAATTAATTAAGTTCAGCCATGATTTCTAAT | 3245 |
| Db | 1501 | ACAGTTCAGTACTAGACAAAACAGCAATTAATTAAGTTCAGCCATGATTTCTAAT | 1560 |
| QY | 3246 | TGGATTAAATTAACCTTAATCACAGTTTAACTCCACAGTGCATTCAATGCAGCTGACAGT | 3305 |
| Db | 1561 | TGGATTAAATTAACCTTAATCACAGTTTAACTCCACAGTGCATTCAATGCAGCTGACAGT | 1620 |
| QY | 3306 | TATATTGTTTTAATGGAGTCAGATATTAATAATCGCTTTGTCACCTCAGGGAGAT | 3365 |
| Db | 1621 | TATATTGTTTTAATGGAGTCAGATATTAATAATCGCTTTGTCACCTCAGGGAGAT | 1680 |
| QY | 3366 | TTACCAATTGTCGGAGACATTTTTTATGATGCATGACATAGGCGAGTTATTTGACATTTAGC | 3425 |
| Db | 1681 | TTACCAATTGTCGGAGACATTTTTTATGATGCATGACATAGGCGAGTTATTTGACATTTAGC | 1740 |
| QY | 3426 | AGTAGAGCCATGGATCTCTCTAATTAATTAACCTGCAATTTGACAGCGCCCAACAAGAAT | 3485 |
| Db | 1741 | AGTAGAGCCATGGATCTCTCTAATTAATTAACCTGCAATTTGACAGCGCCCAACAAGAAT | 1800 |
| QY | 3486 | TATCTCGCCCGAAATGGTATGTCGCAAGGCGATGATACCTTGTTGTTAAAGTTAACTCG | 3545 |
| Db | 1801 | TATCTCGCCCGAAATGGTATGTCGCAAGGCGATGATACCTTGTTGTTAAAGTTAACTCG | 1860 |
| QY | 3546 | TGGCAGACTAGTTCACAGATTTTCCGTGTTGCTGCACAGTATCAATGTTTGAAAAAATTT | 3605 |
| Db | 1861 | TGGCAGACTAGTTCACAGATTTTCCGTGTTGCTGCACAGTATCAATGTTTGAAAAAATTT | 1920 |
| QY | 3606 | TGGCTATTTAAAGATATGATTAATTAAGATGATGCTTTATCTGATTTATTAACCTGATACAATTGA | 3665 |
| Db | 1921 | TGGCTATTTAAAGATATGATTAATTAAGATGATGCTTTATCTGATTTATTAACCTGATACAATTGA | 1980 |
| QY | 3666 | TCTTTTCTAATATTTTCAGAAAGTGATGGGATTAACCTTAGAGAGACTCAAGATATGAT | 3725 |
| Db | 1981 | TCTTTTCTAATATTTTCAGAAAGTGATGGGATTAACCTTAGAGAGACTCAAGATATGAT | 2040 |
| QY | 3726 | TTATATTTAAGTAGAGCTTAAACCCCTTAATTTTCAACAAGTTATATGGCTAAATTT | 3785 |
| Db | 2041 | TTATATTTAAGTAGAGCTTAAACCCCTTAATTTTCAACAAGTTATATGGCTAAATTT | 2100 |
| QY | 3786 | CAGATTGAACAGGGATTCAAGCTATTCGCATCTCATAGAAAGAGAGGCTCCCTCAATC | 3845 |
| Db | 2101 | CAGATTGAACAGGGATTCAAGCTATTCGCATCTCATAGAAAGAGAGGCTCCCTCAATC | 2160 |
| QY | 3846 | TGAAGCTCTCTGAAATCTACCTTCGCAAGCTTCACACAAATCACTGATCTCCCTGAGC | 3905 |
| Db | 2161 | TGAAGCTCTCTGAAATCTACCTTCGCAAGCTTCACACAAATCACTGATCTCCCTGAGC | 2220 |
| QY | 3906 | CACAGGCGCTCATCTCTGAGAGGAGGAAAGATTAGCCAAAGATTAATTTTCATTTCA | 3965 |
| Db | 2221 | CACAGGCGCTCATCTCTGAGAGGAGGAAAGATTAGCCAAAGATTAATTTTCATTTCA | 2280 |
| QY | 3966 | ATCACTTAAGCTGTTAAGACTGATCTGTTTGTAGCAGTTGTTTGTCTCATTTTGGCTGTG | 4025 |
| Db | 2281 | ATCACTTAAGCTGTTAAGACTGATCTGTTTGTAGCAGTTGTTTGTCTCATTTTGGCTGTG | 2340 |
| QY | 4026 | CATTTTGGAGCACTTTGAGAAATATTCATATGSGRSCCTACGATATTTTCTTTTT | 4085 |
| Db | 2341 | CATTTTGGAGCACTTTGAGAAATATTCATATGSGRSCCTACGATATTTTCTTTTT | 2400 |
| QY | 4086 | AATATCTACCTGATATCTTCTCTTAAATTTCTTCAATATGTTTGGCTGATACAC | 4145 |
| Db | 2401 | AATATCTACCTGATATCTTCTCTTAAATTTCTTCAATATGTTTGGCTGATACAC | 2460 |
| QY | 4146 | TGATTTTATTAAGAAATTTAAGAAATCTAACAGCTAAAACTCAGTAAGTACATTAAT | 4205 |
| Db | 2461 | TGATTTTATTAAGAAATTTAAGAAATCTAACAGCTAAAACTCAGTAAGTACATTAAT | 2520 |
| QY | 4206 | TCTCTTAATACATAGACCCGTTGCTACTCTCAGACCCCTCTCTCAATTTTTTTTCTGTGA | 4265 |
| Db | 2521 | TCTCTTAATACATAGACCCGTTGCTACTCTCAGACCCCTCTCTCAATTTTTTTTCTGTGA | 2580 |

QY 4266 GCATGTCATGCTGATTAACCTCATTTTCCTTTATTTCTAATATGGAACATG 4325
|||||
DB 2581 GCATGTCATGCTGATTAACCTCATTTTCCTTTATTTCTAATATGGAACATG 2640
QY 4326 AGGTGACCTCTAATATGAGTGTGTATATAAACCTCTTAGCCATATATAGAAA 4385
|||||
DB 2641 AGGTGACCTCTAATATGAGTGTGTATATAAACCTCTTAGCCATATATAGAAA 2700
QY 4386 TGTCTAATTAAGTACGACATGAAACATGAAATGCTAGTCATGTCCTTTGTCAG 4445
|||||
DB 2701 TGTCTAATTAAGTACGACATGAAACATGAAATGCTAGTCATGTCCTTTGTCAG 2760
QY 4446 CAATTTTGACATGTCATTAATGTTTGTCTAATTTTAAATAAAGTGTCTGGCTTTCAGAA 4505
|||||
DB 2761 CAATTTTGACATGTCATTAATGTTTGTCTAATTTTAAATAAAGTGTCTGGCTTTCAGAA 2820

RESULT 3
ABV21967
ID ABV21967 standard; cDNA; 2905 BP.
XX
AC ABV21967;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 21958.
XX
XX Human prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KM pharmacogenomic marker; gene: ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI Schlegel R, Endege WO, Monahan JE.
XX
XX WPI; 2001-662795/76.
DR
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX Claim 1; Page 3747; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient.
XX (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX

SQ Sequence 2905 BP; 921 A; 509 C; 554 G; 919 T; 2 other;
Query Match 56.1%; Score 2540.4; DB 23; Length 2905;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 2627; Conservative 1; Mismatches 62; Indels 10; Gaps 4;
QY 1664 CTGTGAGATTTGGTCTGATGTATGTACTAAGCAAAATGTTACTGAAATTTCTATCATAA 1723
| | | | |
DB 18 CCGTGCAGATTTGGTCTGATGTATGTACTAAGCAAAATGTTACTGAAATTTCTATCATAA 77
QY 1724 GATGTACAAAGAGCGGAGAACCCAGTATCTTATGCGGAATTTTGGAAACCAAGATCT 1783
| | | | |
DB 78 GATGTACAAAGAGCGGAGAACCCAGTATCTTATGCGGAATTTTGGAAACCAAGATCT 137
QY 1784 CCTAAATTTATTCAGCTCAACAGATTTTCATATCCAGTGAATATTAACATGCATCCAGAGA 1843
| | | | |
DB 138 CCTAAATTTATTCAGCTCAACAGATTTTCATATCCAGTGAATATTAACATGCATCCAGAGA 197
QY 1844 AGCAGAAGAAATTTTAAGTGGGAAATTTATTAAGACCTCATCTTGTATTTCTAGTGTCT 1903
| | | | |
DB 198 AGCAGAAGAAATTTTAAGTGGGAAATTTATTAAGACCTCATCTTGTATTTCTAGTGTCT 257
QY 1904 AGTATTTGGGACTATTTAGTCCAAACCATGAAGACGCAAAAGAAATTTTGTAGACGAG 1963
| | | | |
DB 258 AGTATTTGGGACTATTTAGTCCAAACCATGAAGACGCAAAAGAAATTTTGTAGACGAG 317
QY 1964 AAATCTACTTAAAGAGATTTATTCATCTGATGCAATTTTATGGAAGAAATGTTTGTCTACT 2023
| | | | |
DB 318 AAATCTACTTAAAGAGATTTATTCATCTGATGCAATTTTATGGAAGAAATGTTTGTCTACT 377
QY 2024 GTCAACCAAAATTTATGTCGAAGTCTTCCAGCCCTGCTGCTTGTCCAGACACACAGAGCAA 2083
| | | | |
DB 378 GTCAACCAAAATTTATGTCGAAGTCTTCCAGCCCTGCTGCTTGTCCAGACACACAGAGCAA 437
QY 2084 AATAGAGAGCATCCCACTAGCTAGCACATGACAAAGACATGTTCAATTAATTAACAGA 2143
| | | | |
DB 438 AATAGAGAGCATCCCACTAGCTAGCACATGACAAAGACATGTTCAATTAATTAACAGA 497
QY 2144 TGCCTACTGTGAATATGTTTCCGGAATCACTGTGGAATCTTCCAGTTATTTTCAGACT 2203
| | | | |
DB 498 TGCCTACTGTGAATATGTTTCCGGAATCACTGTGGAATCTTCCAGTTATTTTCAGACT 557
QY 2204 TCAGAAACCTTATTTGATTTTGTTCAGTGTGTCAGTGTAAATCTTCAATTAATTAAGC 2263
| | | | |
DB 558 TCAGAAACCTTATTTGATTTTGTTCAGTGTGTCAGTGTAAATCTTCAATTAATTAAGC 617
QY 2264 AATATTGACATGCTGAAGCAGAAATCACTGATTCATTTACTCCATGCTGTTAAATCT 2323
| | | | |
DB 618 AATATTGACATGCTGAAGCAGAAATCACTGATTCATTTACTCCATGCTGTTAAATCT 677
QY 2324 AAAGAATACTCTCAGTGGGAGAGAAATCTTCCGGGCAATTTTGAATCTCTGCTCCCT 2383
| | | | |
DB 678 AAAGAATACTCTCAGTGGGAGAGAAATCTTGAAGGCAATTTTGAATCTCTGCTCCCT 737
QY 2384 TCCTCTTCTTGTGTTTGTGTCATCTGATTCAGTGTGCGCAAGTATTTGCAATTTCTTCAGA 2443
| | | | |
DB 738 TCCTCTTCTTGTGTTTGTGTCATCTGATTCAGTGTGCGCAAGTATTTGCAATTTCTTCAGA 797
QY 2444 CCAGGCTATTAATGAAGAAACCTTGATTTGTCGTAAGAAATTTGAAGCAGACTAGA 2503
| | | | |
DB 798 CCAGGCTATTAATGAAGAAACCTTGATTTGTCGTAAGAAATTTGAAGCAGACTAGA 857
QY 2504 AAATCATATACAAATTTTACCTGCTCAAGATGGAACCTCCTCTCCAGCTTATGATTT 2563
| | | | |
DB 858 AAATCATATACAAATTTTACCTGCTCAAGATGGAACCTCCTCTCCAGCTTATGATTT 917
QY 2564 TCTAATGATGATGATGTCGCAACATCTCAACGTGCACTAGAAAGTTCCCAAGTGTAT 2623
| | | | |
DB 918 TCTAATGATGATGATGTCGCAACATCTCAACGTGCACTAGAAAGTTCCCAAGTGTAT 977
QY 2624 GAAAGAAACAGATGTCAGAGAAATGATTAAGAAACAATGAAGATTAATTCGCAAGTCAG 2683
| | | | |
DB 978 GAAAGAAACAGATGTCAGAGAAATGATTAAGAAACAATGAAGATTAATTCGCAAGTCAG 1037

OY 2684 AAAAGAACCGATTGAAACTCTGGAATTAAGCATTTGGATAGAGTAATTTGGTTAAAGA 2743
 DB 1038 AAAAGAACCGATTGAAACTCTGGAATTAAGCATTTGGATAGAGTAATTTGGTTAAAGA 1097
 OY 2744 AACGAAAAATCATTTAGACGTGATTAAGAGTTAGAGTCTCAAAAAGTAACTAATTTTA 2803
 DB 1098 AACGAAAAATCATTTAGACGTGATTAAGAGTTAGAGTCTCAAAAAGTAACTAATTTTA 1157
 OY 2804 TAGGGCTGTGGTTCCAAAATTTTGGCATGATAGACTTAATTTATTTCTTAAAGAA 2863
 DB 1158 TAGGGCTGTGGTTCCAAAATTTTGGCATGATAGACTTAATTTATTTCTTAAAGAA 1217
 OY 2864 TAATATTAATCATTTCAAGTTGACAGACTAGTCCATCCATAGATATATATATAGT 2923
 DB 1218 TAATATTAATCATTTCAAGTTGACAGACTAGTCCATCCATAGATATATATATAGT 1277
 OY 2924 CACATATTTTATTTAAATTTTCTAGTAACATTAATAACAAAGTAAAGTGAGCAGGCG 2983
 DB 1278 CACATATTTTATTTAAATTTTCTAGTAACATTAATAACAAAGTAAAGTGAGCAGGCG 1337
 OY 2984 AAAATATATTTGATATATTTTACCTTTTACCCAGTATACCCAAATATGCAATATAGAA 3043
 DB 1338 AAAATATATTTGATATATTTTACCTTTTACCCAGTATACCCAAATATGCAATATAGAA 1397
 OY 3044 TTATTAATGAGATATTTTACATCTTTTGTACCAAGTCTCTAAATGACATATAT 3103
 DB 1398 TTATTAATGAGATATTTTACATCTTTTGTACCAAGTCTCTAAATGACATATAT 1457
 OY 3104 TTATATCTAGTCAATTTCTTACTCCGAGTAGCCATATTTCAAGTGTTCATGCCACATG 3163
 DB 1458 TTATATCTAGTCAATTTCTTACTCCGAGTAGCCATATTTCAAGTGTTCATGCCACATG 1517
 OY 3164 TGGCCGTGACACGTGATTTGACAGTACAGTACAGAAAAAATCTAGCATATATAT 3223
 DB 1518 TGGCCGTGACACGTGATTTGACAGTACAGTACAGAAAAAATCTAGCATATATAT 1577
 OY 3224 AGTTCTAGCATGATTTCTATTTGATTAATAATTAACCTATACAGTAACTCCACA 3283
 DB 1578 AGTTCTAGCATGATTTCTATTTGATTAATAATTAACCTATACAGTAACTCCACA 1637
 OY 3284 GTGCATTCATGACAGTACAGTATATTTGTTTATTTGAGTCAATGATTAATAATCAGC 3343
 DB 1638 GTGCATTCATGACAGTACAGTATATTTGTTTATTTGAGTCAATGATTAATAATCAGC 1697
 OY 3344 GTTGTCAACCTCAGGGGATTTAGCAATGTGCGGAGACATTTTGTATGTCATGACATA 3403
 DB 1698 GTTGTCAACCTCAGGGGATTTAGCAATGTGCGGAGACATTTTGTATGTCATGACATA 1757
 OY 3404 GGGCAGTTATGACATTTAGTAGTAGAGGCCATGATCCTGCTAAATTAACCTGCATTTGG 3463
 DB 1758 GGGCAGTTATGACATTTAGTAGTAGAGGCCATGATCCTGCTAAATTAACCTGCATTTGG 1817
 OY 3464 ACAGCGCCCCACACAAAGATTAATCTGCCCCGAAATGAGTAGTCTGCCAAGCGTAGATA 3523
 DB 1818 ACAGCGCCCCACACAAAGATTAATCTGCCCCGAAATGAGTAGTCTGCCAAGCGTAGATA 1877
 OY 3524 ACCTGTGTTAAAGTAACCTGCGCAGACTAGTTCACAGATTTCCGAGTTCGTCGCTCA 3583
 DB 1878 ACCTGTGTTAAAGTAACCTGCGCAGACTAGTTCACAGATTTCCGAGTTCGTCGCTCA 1937
 OY 3584 CGTATCATGTTTGAATAATTTGGCTATTAAGATATGATATAGATGCTTATATCTGA 3643
 DB 1938 CGTATCATGTTTGAATAATTTGGCTATTAAGATATGATATAGATGCTTATATCTGA 1997
 OY 3644 TTATTAATGATGATACACTGATCTTTCTAATTAATTTTCAAGAGTATGATGATACCT 3703
 DB 1998 TTATTAATGATGATACACTGATCTTTCTAATTAATTTTCAAGAGTATGATGATACCT 2057
 OY 3704 AGAAGAGGACCTCAGATGATATTTATTTTAAGAGCTTAATAACCTCCTTATATTC 3763
 DB 2058 AGAAGAGGACCTCAGATGATATTTATTTTAAGAGCTTAATAACCTCCTTATATTC 2117

OY 3764 TACAAGTTATATGCTAAATTTACAGATTAAGACAGGATTCAGCATTCGCCATCTCCTCA 3823
 DB 2118 TACAAGTTATATGCTAAATTTACAGATTAAGACAGGATTCAGCATTCGCCATCTCCTCA 2177
 OY 3824 TGAAGAGAGGCTCCCTCATCTGAAGCGTCTCTGAATCTACCTTTGCAAGCTTCAGAC 3883
 DB 2178 TGAAGAGAGGCTCCCTCATCTGAAGCGTCTCTGAATCTACCTTTGCAAGCTTCAGAC 2237
 OY 3884 AATACGATGATCTCCCTCAGCACACGCGCTCATCTCTGTAGGAGGAGGAAGATTAAGC 3943
 DB 2238 AATACGATGATCTCCCTCAGCACACGCGCTCATCTCTGTAGGAGGAGGAAGATTAAGC 2297
 OY 3944 AAGAGTTAATTTTCAATCTCAATCAATCAATAGCTGTAGACTATCTGTAGAGCTTG 4003
 DB 2298 AAGAGTTAATTTTCAATCTCAATCAATCAATAGCTGTAGACTATCTGTAGAGCTTG 2357
 OY 4004 TTGTCTCATTTTGTGCTGTGCATTTTGTGACATTTTGTGAGANATTTCTATTTGGT 4063
 DB 2358 TTGTCTCATTTTGTGCTGTGCATTTTGTGAGACATTTTGTGAGANATTTCTATTTGGT 2417
 OY 4064 GCTCTACGATTTTCTTTTAATATCTACATGATATCTGTCTTTAATTTCTCA 4123
 DB 2418 GCTCTACGATTTTCTTTTAATATCTACATGATATCTGTCTTTAATTTCTCA 2477
 OY 4124 CATATGTTTGGCTGATACACATGATTTTATTAAGTAAAGATTAAGATTAAGAGTA 4183
 DB 2478 CATATGTTTGGCTGATACACATGATTTTATTAAGTAAAGATTAAGATTAAGAGTA 2536
 OY 4184 AAATCAGTAACTGATTTTCTTTCTTATTAACATGACACCGTTGCTACTCTCAGACCT 4243
 DB 2537 AAATCAGTAACTGATTTTCTTTCTTATTAACATGACACCGTTGCTACTCTCAGACCT 2592
 OY 4244 CTCTCTCAATTTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 4303
 DB 2593 CTCTCTCAATTTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 2647
 OY 4304 TATTTCTAATATGAGAACATGAGAGTGAACCTTAATTAAGTGTAGTAAATTAACAT 4363
 DB 2648 TATTTCTAATATGAGAACATGAGAGTGAACCTTAATTAAGTGTAGTAAATTAACAT 2707

RESULT 4
 ABV27799
 ID ABV27799 standard; cDNA; 2905 BP.
 XX
 AC ABV27799;
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 27790.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200160860-A2.
 PD
 XX 23-AUG-2001.
 PF
 XX 20-FEB-2001; 2001WO-US05171.
 PR 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 BA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;

DR WPI: 2001-662795/76.
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1: Page 5714-5715; 11750bp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 2905 BP; 921 A; 509 C; 554 G; 919 T; 2 other;
Query Match 56.1%; Score 2540.4; DB 23; Length 2905;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 2627; Conservative 1; Mismatches 62; Indels 10; Gaps 4;
QY 1664 CTGTGAGATTGGTCTGATGTATGTACTAAGCAAAATGTTACTGAATTTCCATCATATAA 1723
DB 18 CCGGAGATTGGTCTGATGTATGTACTAAGCAAAATGTTACTGAATTTCCATCATATAA 77
QY 1724 GAGGTACAAAGAGCCAGAACCCAGATATCTTATGCTGGAATGTAGAACCAAAATCT 1783
DB 78 GAGGTACAAAGAGCCAGAACCCAGATATCTTATGCTGGAATGTAGAACCCGAATCT 137
QY 1784 CCTAAATTTATTCAGCTCAACAGATTTATCTATCCAGTAATTAATCATGATCCAGA 1843
DB 138 CTTAAATTTATTCAGCTCAACAGATTTATCTATCCAGTAATTAATCATGATCCAGA 197
QY 1844 AGCAGAAATATTTAAGTGGGGAATTAATAAGACCTCATCTTGTAATCTACTGTCTC 1903
DB 198 AGCAGAAATATTTAAGTGGGGAATTAATAAGACCTCATCTTGTAATCTACTGTCTC 257
QY 1904 AGATTGGGACTATTTAGTCCCAACATGAAAGCAAGAAAGATTTTGTAGTGAAGCAG 1963
DB 258 AGATTGGGACTATTTAGTCCCAACATGAAAGCAAGAAAGATTTTGTAGTGAAGCAG 317
QY 1964 AAATCTACCTAAAGAGATGTATCTACTGGAATTTATTTGAAGAAGATGTTTGTACT 2023
DB 318 AAATCTACCTAAAGAGATGTATCTACTGGAATTTATTTGAAGAAGATGTTTGTACT 377
QY 2024 GTCAACCAAAATATGCTGCAAGTCTTCCAGCCCTGCTGCTTGCAGACACAGAAAGCAA 2083
DB 378 GTCAACCAAAATATGCTGCAAGTCTTCCAGCCCTGCTGCTTGCAGACACAGAAAGCAA 437
QY 2084 AATAGAGAGATCCCATCTACTAGACACATGCAAGACATAGTTCGAATTAATAACAGA 2143
DB 438 AATAGAGAGATCCCATCTACTAGACACATGCAAGACATAGTTCGAATTAATAACAGA 497
QY 2144 TGCACCTCTGGAATGTTTCCGGAATCAGTGTGAAATCTTCCAGTTATTTCACT 2203
DB 498 TGCACCTCTGGAATGTTTCCGGAATCAGTGTGAAATCTTCCAGTTATTTCACT 557
QY 2204 TCAGAAACCAATATGATTTTGTTCAGTGTGCACTGTAATCTCAATATAAAAAAGC 2263
DB 558 TCAGAAACCAATATGATTTTGTTCAGTGTGCACTGTAATCTCAATATAAAAAAGC 617
QY 2264 AATATTACACTGTGTAAGAGAAATCTTGATTTCACTTACTCATGCTGTTAAATCT 2323
DB 618 AATATTACACTGTGTAAGAGAAATCTTGATTTCACTTACTCATGCTGTTAAATCT 677

QY 2324 AAAGAAATCTCCAGTGGGAGAGAAATCTTCCGGCATATTTTGATCTCTGCCCTCCCT 2383
DB 678 AAAGAAATCTCCAGTGGGAGAGAAATCTTCCGGCATATTTTGATCTCTGCCCTCCCT 737
QY 2384 TCCCTCTCTGTTTGTGTGATCTGCAATTCAGGTGGCCAGATATTTCCATCTCTTACA 2443
DB 738 TCCCTCTCTGTTTGTGTGATCTGCAATTCAGGTGGCCAGATATTTCCATCTCTTACA 797
QY 2444 CCAGGCTATATTAAGAAACCTTGATTTGTGGCTCAGAAATTAAGACAGACTAGA 2503
DB 798 CCAGGCTATATTAAGAAACCTTGATTTGTGGCTCAGAAATTAAGACAGACTAGA 857
QY 2504 AAATCATATCACAATTTTACCTGCTCAAGATGGAACCTCTCTTCCAGCTTATGATTT 2563
DB 858 AAATCATATCACAATTTTACCTGCTCAAGATGGAACCTCTCTTCCAGCTTATGATTT 917
QY 2564 TCTAATGATATGATGATCCCAACATCTCAACCTGACATGAGAAAGTCCCAAGTAT 2623
DB 918 TCTAATGATATGATGATCCCAACATCTCAACCTGACATGAGAAAGTCCCAAGTAT 977
QY 2624 GAAAGAAACAGATGTGAGAGAAATGATTAAGAAACATGAAGATTAATTCGGCAGTCA 2683
DB 978 GAAAGAAACAGATGTGAGAGAAATGATTAAGAAACATGAAGATTAATTCGGCAGTCA 1037
QY 2684 AAAAGAACCGATTGAACCTCTGAGAAATTAAGCATTTGGAATAGAATTTGTTTAAAGA 2743
DB 1038 AAAAGAACCGATTGAACCTCTGAGAAATTAAGCATTTGGAATAGAATTTGTTTAAAGA 1097
QY 2744 AGCAGAAATATCATTTTACAGCTGATTAAGAGTTTATGATGCTCAAAAGTGAATTTTA 2803
DB 1098 AGCAGAAATATCATTTTACAGCTGATTAAGAGTTTATGATGCTCAAAAGTGAATTTTA 1157
QY 2804 TAGGGCTGTGGTTCCAAAATTTTGGCATGATAGACTTAATTTTCCCTTAAAGAA 2863
DB 1158 TAGGGCTGTGGTTCCAAAATTTTGGCATGATAGACTTAATTTTCCCTTAAAGAA 1217
QY 2864 TAATATTAATCATATTTTCAAGTTTGCAGACTAGTCCATCAATTAAGATTAATATAGT 2923
DB 1218 TAATATTAATCATATTTTCAAGTTTGCAGACTAGTCCATCAATTAAGATTAATATAGT 1277
QY 2924 CACATATTTTATTTAAATTTTCTAGTACTACTTTAAACAAAGTAAAGTGGACAGGCG 2983
DB 1278 CACATATTTTATTTAAATTTTCTAGTACTACTTTAAACAAAGTAAAGTGGACAGGCG 1337
QY 2984 AAAATATTTTGAATATCTTCCAGCAGTATGATCCCAAAATAGCAGAAATTAAGAA 3043
DB 1338 AAAATATTTTGAATATCTTCCAGCAGTATGATCCCAAAATAGCAGAAATTAAGAA 1397
QY 3044 TTATTAATGAGATATTTTACATCTTTTGTATCAAGTCTTCTAAATGAGATACATATT 3103
DB 1398 TTATTAATGAGATATTTTACATCTTTTGTATCAAGTCTTCTAAATGAGATACATATT 1457
QY 3104 TTATTAATGAGATATTTTACATCTTTTGTATCAAGTCTTCTAAATGAGATACATATT 3163
DB 1458 TTATTAATGAGATATTTTACATCTTTTGTATCAAGTCTTCTAAATGAGATACATATT 1517
QY 3164 TGGCCTGTGACTGATGATTTGAGACAGTTCACTAGTACACAAAACCTGCAATTAATCT 3223
DB 1518 TGGCCTGTGACTGATGATTTGAGACAGTTCACTAGTACACAAAACCTGCAATTAATCT 1577
QY 3224 AGTTTACGACATGATTTCTATTTGGATTAATAATTAATCTAATCAGAGTTAACTCACA 3283
DB 1578 AGTTTACGACATGATTTCTATTTGGATTAATAATTAATCTAATCAGAGTTAACTCACA 1637
QY 3284 GTGCATTCATGACGCTACAGATTAATTTGTTTATTTGAGAGTCAATGATTAATAATCAGC 3343
DB 1638 GTGCATTCATGACGCTACAGATTAATTTGTTTATTTGAGAGTCAATGATTAATAATCAGC 1697
QY 3344 GTTGTCAACCTCAGGGGATTTAGCAATTTGCGGAGACATTTTGTGATGTCAGTACTA 3403
DB 1698 GTTGTCAACCTCAGGGGATTTAGCAATTTGCGGAGACATTTTGTGATGTCAGTACTA 1757


```
QY 1866 GAATTATATAAGACCTCATCTGTGTATCTAGTGTGTGACAGTATGGACATATTTAGTCCA 1925
DB 181 GAAATTATATAAGACCTCATCTGTGTATCTAGTGTGTGACAGTATGGACATATTTAGTCCA 240
QY 1926 ACCATGAAACAGCAAGAAAGATTTAGTGAAGCAGCAAACTCCATAAAGGATATGTT 1985
DB 241 ACCATGAAACAGCAAGAAAGATTTAGTGAAGCAGCAAACTCCATAAAGGATATGTT 300
QY 1986 ATCAGTGAATTTATTTCTGAAGAAGATTTTGTCTACTGTCAACCAATATGTGTCAAGT 2045
DB 301 ATCAGTGAATTTATTTCTGAAGAAGATTTTGTCTACTGTCAACCAATATGTGTCAAGT 360
QY 2046 CTTCACACCTCTGCTGCTGCCAGACACACAGAAAGCAAAATAGAGACATCCACTAGCT 2105
DB 361 CTTCACACCTCTGCTGCTGCCAGACACACAGAAAGCAAAATAGAGACATCCACTAGCT 420
QY 2106 AGCACACATGCACAGACATAGTTCAATATATACAGATGCATCTAGTGAATGTTCCG 2165
DB 421 AGCACACATGCACAGACATAGTTCAATATATACAGATGCATCTAGTGAATGTTCCG 480
QY 2166 GAAATCTGTGGAATAATCTTCCAGTATTTTTCAGACTTCAGAAACCATTTATGATTTG 2225
DB 481 GAAATCTGTGGAATAATCTTCCAGTATTTTTCAGACTTCAGAAACCATTTATGATTTG 540
QY 2226 TTCACTGATGGCACTGTAAATCTCAATATATATAAAGCAATATTGACACTGTGAAGCAG 2285
DB 541 TTCACTGATGGCACTGTAAATCTCAATATATAAAGCAATATTGACACTGTGAAGCAG 600
QY 2286 AATATCTGTGATCATTTACTCCATGCTGTGTTAAATCTAATAAGAAATCTCCAGTGGGAGA 2345
DB 601 AATATCTGTGATCATTTACTCCATGCTGTGTTAAATCTAATAAGAAATCTCCAGTGGGAGA 660
QY 2346 GGAATCTGTGGGCATATTTTGAATCTCTGCTCCCTCTCTCTCTCTCTCTCTCTCTCTCT 2405
DB 661 GGAATCTGTGGGCATATTTTGAATCTCTGCTCCCTCTCTCTCTCTCTCTCTCTCTCTCT 720
QY 2406 CTGCATTCAGTGGCCAGATTTTGCATTTCTTCCTTCACACAGGCTATATTTGAAGAAAC 2465
DB 721 CTGCATTCAGTGGCCAGATTTTGCATTTCTTCCTTCACACAGGCTATATTTGAAGAAAC 780
QY 2466 CTGTG 2469
DB 781 CTGTG 784

RESULT 6
ABV12212
ID ABV12212 standard; cDNA; 360 BP.
XX
AC ABV12212;
XX
AC 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 12203.
XX
KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW Pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN MO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
```

```
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer
PS Claim 1; Page 2006; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (1) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (1) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 360 BP; 124 A; 68 C; 73 G; 94 T; 1 other;
XX
Query Match 7.7%; Score 349.8; DB 23; Length 360;
Best Local Similarity 99.4%; Pred. No. 2.5e-71;
Matches 351; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1725 ATGACACAGAAAGCGGAGAACCCAGTATCTTATGCGAATGTTAGAAACCAAGATCTC 1784
DB 7 AGGTACAGAAAGCGGAGAACCCAGTATCTTATGCGAATGTTAGAAACCAAGATCTC 66
QY 1785 CTAAATTTATCCAGCTCAACAGAGATTTCAATTCAGTGAATTAACATGATCAAGAA 1844
DB 67 CTAAATTTATCCAGCTCAACAGAGATTTCAATTCAGTGAATTAACATGATCAAGAA 126
QY 1845 GCAGAGAAATTTAAGTGGGGAATTTATTAAGACCTCTCTTGTATCTATGTGTCA 1904
DB 127 GCAGAGAAATTTAAGTGGGGAATTTATTAAGACCTCTCTTGTATCTATGTGTCA 186
QY 1905 GATTTGGACTATTTAGTCCAAACCATGAAGAAACGAAAGATTTAGTGAAGCAGGA 1964
DB 187 GATTTGGACTATTTAGTCCAAACCATGAAGAAACGAAAGATTTAGTGAAGCAGGA 246
QY 1965 AACTACCTAAAGAGATTTATCTACTGCAATTTATTTGAGAGAGATGTTTGTACTG 2024
DB 247 AACTACCTAAAGAGATTTATCTACTGCAATTTATTTGAGAGAGATGTTTGTACTG 306
QY 2025 TCAACCAAAATATGCTGCAAGTCTTCAGCCCTCTCTCTCCAGACACAGA 2077
DB 307 TCAACCAAAATATGCTGCAAGTCTTCAGCCCTCTCTCTCCAGACACAGA 359

RESULT 7
ABV03043
ID ABV03043 standard; cDNA; 374 BP.
XX
AC ABV03043;
XX
AC 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 3034.
XX
KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW Pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
```



```

XX XX WO200160860-A2.
XX PN
XX 23-AUG-2001.
XX PD
XX 20-FEB-2001; 2001WO-US05171.
XX PF
XX 17-FEB-2000; 2000US-183119P.
XX PR 16-MAR-2000; 2000US-189862P.
XX PR 25-MAY-2000; 2000US-207454P.
XX PR 09-JUN-2000; 2000US-211314P.
XX PR 18-JUL-2000; 2000US-219007P.
XX PR 13-DEC-2000; 2000US-255281P.
XX XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Schlegel R, Endege WO, Monahan JE;
XX DR WPI; 2001-662795/76.
XX XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
XX PT prostate cells and correlating with presence of prostate cancer, useful
XX PR for detecting presence of prostate cancer, stage of prostate cancer -
XX PS
XX PS Claim 1; Page 561; 11750pp; English.
XX XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX CC specification or its complement. (I) is useful for:
XX CC (a) assessing whether a patient is afflicted with prostate cancer;
XX CC (b) monitoring the progression of prostate cancer in a patient;
XX CC (c) assessing the efficacy of a test compound to inhibit prostate
XX CC cancer in a patient;
XX CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX CC in a patient;
XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;
XX CC (f) assessing the prostate cell carcinogenic potential of a compound;
XX CC (g) determining whether prostate cancer has metastasized in a patient;
XX CC (h) assessing the aggressiveness or indolence of prostate cancer in a
XX CC patient;
XX CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX XX
XX Sequence 374 BP; 123 A; 75 C; 81 G; 95 T; 0 other;
SQ
Query Match 7.7%; Score 348.8; DB 23; Length 374;
Best Local Similarity 99.4%; Pred. No. 4.3e-71;
Matches 350; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1725 ATGTACAAGAAAGGCGAGAACCCAGTATCTTATGCTGGAATGTAGAACCAAGATCTC 1784
DB 23 AGGTACAAGAAAGGCGAGAACCCAGTATCTTATGCTGGAATGTAGAACCAAGATCTC 82
QY 1785 CTAATAATTTATCCAGCTCAACAGATTCATATCCAGTGAATATTAACATGCATCCAGAA 1844
DB 83 CTAATAATTTATCCAGCTCAACAGATTCATATCCAGTGAATATTAACATGCATCCAGAA 142
QY 1845 GCAGAGAATATTTAAGTGGGGAATTTATATAAAGACCTCATCTTGATCTGTGTGTCA 1904
DB 143 GCAGAGAATATTTAAGTGGGGAATTTATATAAAGACCTCATCTTGATCTGTGTGTCA 202
QY 1905 GATATGGAGCTATTTAGTCCACATGAAACAGCAAAAGAGATTTTGTAGACAGGA 1964
DB 203 GATATGGAGCTATTTAGTCCACATGAAACAGCAAAAGAGATTTTGTAGACAGGA 262
QY 1965 AACTACCTAAAGAGATATGTTATTCACCTGGAATTTATCTGGAAGAGATGTTTGTACTG 2024
DB 263 AACTACCTAAAGAGATATGTTATTCACCTGGAATTTATCTGGAAGAGATGTTTGTACTG 322
QY 2025 TCAACCAATATGTCGCAAGCTTCCAGCCCTGCTGCTGTCGACAGACACAG 2076
DB 323 TCAACCAATATGTCGCAAGCTTCCAGCCCTGCTGCTGTCGACAGACACAG 374

```

```

RESULT 8
ABV33357
ID ABV33357 standard; cDNA; 397 BP.
XX AC
XX ABV33357;
XX DT
XX 16-SEP-2002 (first entry)
XX DE
XX Human prostate expression marker cDNA 33348.
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KW pharmacogenomic marker; gene; ss.
XX OS Homo sapiens.
XX XX
XX WO200160860-A2.
XX PN
XX 23-AUG-2001.
XX PD
XX 20-FEB-2001; 2001WO-US05171.
XX PF
XX 17-FEB-2000; 2000US-183119P.
XX PR 16-MAR-2000; 2000US-189862P.
XX PR 25-MAY-2000; 2000US-207454P.
XX PR 09-JUN-2000; 2000US-211314P.
XX PR 18-JUL-2000; 2000US-219007P.
XX PR 13-DEC-2000; 2000US-255281P.
XX XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Schlegel R, Endege WO, Monahan JE;
XX DR WPI; 2001-662795/76.
XX XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
XX PT prostate cells and correlating with presence of prostate cancer, useful
XX PR for detecting presence of prostate cancer, stage of prostate cancer -
XX PS
XX PS Claim 1; Page 7066; 11750pp; English.
XX XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX CC specification or its complement. (I) is useful for:
XX CC (a) assessing whether a patient is afflicted with prostate cancer;
XX CC (b) monitoring the progression of prostate cancer in a patient;
XX CC (c) assessing the efficacy of a test compound to inhibit prostate
XX CC cancer in a patient;
XX CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX CC in a patient;
XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;
XX CC (f) assessing the prostate cell carcinogenic potential of a compound;
XX CC (g) determining whether prostate cancer has metastasized in a patient;
XX CC (h) assessing the aggressiveness or indolence of prostate cancer in a
XX CC patient;
XX CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX XX
XX Sequence 397 BP; 129 A; 80 C; 88 G; 100 T; 0 other;
SQ
Query Match 7.7%; Score 348.8; DB 23; Length 397;
Best Local Similarity 99.4%; Pred. No. 4.4e-71;
Matches 350; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1725 ATGTACAAGAAAGGCGAGAACCCAGTATCTTATGCTGGAATGTAGAACCAAGATCTC 1784
DB 46 AGGTACAAGAAAGGCGAGAACCCAGTATCTTATGCTGGAATGTAGAACCAAGATCTC 105
QY 1785 CTAATAATTTATCCAGCTCAACAGATTCATATCCAGTGAATATTAACATGCATCCAGAA 1844
DB 106 CTAATAATTTATCCAGCTCAACAGATTCATATCCAGTGAATATTAACATGCATCCAGAA 165
QY 1845 GCAGAGAATATTTAAGTGGGGAATTTATATAAAGACCTCATCTTGATCTGTGTGTCA 1904
DB 166 GCAGAGAATATTTAAGTGGGGAATTTATATAAAGACCTCATCTTGATCTGTGTGTCA 225

```

| | | | |
|---------------------------------------|---|---|------|
| QY | 1905 | GTATTGGAGCTATTATTAGTCCACCATGAAACAGCAAAAGAGATTTTAGTGAAGCAGGA | 1964 |
| | | | |
| DB | 226 | GTATTGGAGCTATTATTAGTCCACCATGAAACAGCAAAAGAGATTTTAGTGAAGCAGGA | 285 |
| QY | 1965 | AACTACCTAAAGAGATATGTTATCACTGAAATTTATCTGAAGAAGATGTTTGCTACTG | 2024 |
| | | | |
| DB | 286 | AACTACCTAAAGAGATATGTTATCACTGAAATTTATCTGAAGAAGATGTTTGCTACTG | 345 |
| QY | 2025 | TCAACCAAAATATGCTGCAAGCTCTCCAGCCGCTGCTCCAGACACAG | 2076 |
| | | | |
| DB | 346 | TCAACCAAAATATGCTGCAAGCTCTCCAGCCGCTGCTGCTCCAGACACAG | 397 |
| RESULT 9 | | | |
| | ABV42280 | | |
| ID | ABV42280 | standard; cDNA; 397 BP. | |
| XX | | | |
| AC | ABV42280; | | |
| XX | | | |
| DT | 16-SEP-2002 | (first entry) | |
| XX | | | |
| DE | Human prostate expression marker | CDNA 42271. | |
| XX | | | |
| KW | Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker; | | |
| KW | pharmacogenomic marker; gene; ss. | | |
| OS | Homo sapiens. | | |
| XX | | | |
| PN | WO200160860-A2. | | |
| XX | | | |
| PD | 23-AUG-2001. | | |
| XX | | | |
| PF | 20-FEB-2001; 2001WO-US05171. | | |
| XX | | | |
| PR | 17-FEB-2000; 2000US-183319P. | | |
| PR | 16-MAR-2000; 2000US-189862P. | | |
| PR | 25-MAY-2000; 2000US-207454P. | | |
| PR | 09-JUN-2000; 2000US-211314P. | | |
| PR | 18-JUL-2000; 2000US-219007P. | | |
| PR | 13-DEC-2000; 2000US-255281P. | | |
| XX | | | |
| PA | (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC. | | |
| XX | | | |
| PI | Schlegel R, Endege WO, Monahan JE; | | |
| DR | WPI; 2001-662795/76. | | |
| XX | | | |
| PT | Novel isolated nucleic acid molecule associated with cancerous state of | | |
| PT | prostate cells and correlating with presence of prostate cancer, useful | | |
| PT | for detecting presence of prostate cancer, stage of prostate cancer - | | |
| PS | Claim 1; Page 8470-8471; 11750pp; English. | | |
| XX | | | |
| CC | The invention relates to an isolated nucleic acid molecule (I) comprising | | |
| CC | a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the | | |
| CC | specification or its complement. (I) is useful for: | | |
| CC | (a) assessing whether a patient is afflicted with prostate cancer; | | |
| CC | (b) monitoring the progression of prostate cancer in a patient; | | |
| CC | (c) assessing the efficacy of a test compound to inhibit prostate | | |
| CC | cancer in a patient; | | |
| CC | (d) assessing the efficacy of a therapy for inhibiting prostate cancer | | |
| CC | in a patient; | | |
| CC | (e) selecting a composition for inhibiting prostate cancer in a patient; | | |
| CC | (f) assessing the prostate cell carcinogenic potential of a compound; | | |
| CC | (g) determining whether prostate cancer has metastasized in a patient; | | |
| CC | (h) assessing the aggressiveness or indolence of prostate cancer in a | | |
| CC | patient; | | |
| CC | (I) is also useful as a pharmacodynamic or pharmacogenomic marker. | | |
| XX | | | |
| SQ | Sequence 397 BP; 129 A; 80 C; 88 G; 100 T; 0 other; | | |
| Query Match | | | |
| 7.7%; Score 348.8; DB 23; Length 397; | | | |

| Best Local Similarity | 99.4% | Pred. No. 4.4e-71 | Mismatches | 2 | Indels | 0 | Gaps | 0 |
|-----------------------|--------------|--|------------|----|--------|----|------|----|
| Matches 350; | Conservative | 0; | Mismatches | 2; | Indels | 0; | Gaps | 0; |
| QY | 1725 | ATGTACAAGAAAGCGGAGAACCCAGTATCTTATGCTGCAATGTGAGAACCAAGATCTC | 1784 | | | | | |
| Db | 46 | AGGTCAAGAAAGGCGAGAACCCAGTATCTTATGCTGCAATGTGAGAACCAAGATCTC | 105 | | | | | |
| QY | 1785 | CTAAATTTATTCACCTCAACAGGATTTTCATATCCAGTGAATATTAACATCGATCCAGAA | 1844 | | | | | |
| Db | 106 | CTAAATTTATTCACCTCAACAGGATTTTCATATCCAGTGAATATTAACATCGATCCAGAA | 165 | | | | | |
| QY | 1845 | GCAGAGAATATTTTAAAGTGGGGAATTTATATAAAGACCTCATCTTGTATCTAGTGTC | 1904 | | | | | |
| Db | 166 | GCAGAGAATATTTTAAAGTGGGGAATTTATATAAAGACCTCATCTTGTATCTAGTGTC | 225 | | | | | |
| QY | 1905 | GTATTGGGACATATTTAGTCCAAACCATGAAACAGCAAAAAGAAATTTTATGTAAGCAGA | 1964 | | | | | |
| Db | 226 | GTATTGGGACATATTTAGTCCAAACCATGAAACAGCAAAAAGAAATTTTATGTAAGCAGA | 285 | | | | | |
| QY | 1965 | AACATCGAAAAAGGATATGTTTATCTCACTGGAAATTTTATGGAAGAAGATGTTTGTCTAG | 2024 | | | | | |
| Db | 286 | AACATCGAAAAAGGATATGTTTATCTCACTGGAAATTTTATGGAAGAAGATGTTTGTCTAG | 345 | | | | | |
| QY | 2025 | TCAACCAATATATGTCGAAGTCTTCAGCCCTGCTGCTTCCAGACACACAG | 2076 | | | | | |
| Db | 346 | TCAACCAATATATGTCGAAGTCTTCAGCCCTGCTGCTTCCAGACACACAG | 397 | | | | | |
| RESULT 10 | | | | | | | | |
| AAC28836 | | | | | | | | |
| ID | AAC28836 | standard; cDNA; 355 BP. | | | | | | |
| XX | XX | | | | | | | |
| AC | AC | AAC28836; | | | | | | |
| XX | XX | | | | | | | |
| DT | DT | 06-OCT-2000 (first entry) | | | | | | |
| XX | XX | | | | | | | |
| DE | DE | Human secreted protein 5' EST, SEQ ID NO: 32911. | | | | | | |
| XX | XX | | | | | | | |
| KW | KW | Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation; | | | | | | |
| XX | XX | gene therapy; chromosome mapping; ss. | | | | | | |
| OS | OS | Homo sapiens. | | | | | | |
| XX | XX | | | | | | | |
| PN | PN | EPI033401-A2. | | | | | | |
| PD | PD | 06-SEP-2000. | | | | | | |
| XX | XX | | | | | | | |
| PF | PF | 21-FEB-2000; 2000EP-0200610. | | | | | | |
| XX | XX | | | | | | | |
| PR | PR | 26-FEB-1999; 99US-0122487. | | | | | | |
| PT | PT | (GEST) GENSET. | | | | | | |
| XX | XX | | | | | | | |
| PS | PS | Dumas Milne Edwards J, Duclert A, Giordano J; | | | | | | |
| XX | XX | WPI: 2000-500381/45. | | | | | | |
| CC | CC | The present sequence is one of a large number of 5' ESTs derived from | | | | | | |
| CC | CC | mRNAs encoding secreted proteins. No ORF has yet been conclusively | | | | | | |
| CC | CC | identified within the present sequence. The 5' ESTs were prepared from | | | | | | |
| CC | CC | total human RNAs or polyA+ RNAs derived from 30 different tissues. EST | | | | | | |
| CC | CC | sequences usually correspond mainly to the 3' untranslated region (UTR) | | | | | | |
| CC | CC | of the mRNA because they are often obtained from oligo-dT primed cDNA | | | | | | |
| CC | CC | libraries. Such ESTs are not well suited for isolating cDNA sequences | | | | | | |
| CC | CC | derived from the 5' ends of mRNAs and even in those cases where longer | | | | | | |
| CC | CC | cDNA sequences have been obtained, the full 5' UTR is rarely included. | | | | | | |
| CC | CC | 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be | | | | | | |

PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX
XX Claim 1; Page 5006; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (1) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (1) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX Sequence 416 BP; 109 A; 67 C; 57 G; 159 T; 24 other;
SQ
Query Match 6.7%; Score 304.4; DB 23; Length 416;
Best Local Similarity 87.1%; Pred. No. 9, 4e-61;
Matches 317; Conservative 1; Mismatches 46; Indels 0; Gaps 0;
QY 4090 TCTACTGATATCTGTTCTTTAAATTTCTTCACATATGTTGGCTGATACACTGAT 4149
DB 21 TCTACTGATATCTGTTCTTTAAATTTCTTCACATATGTTGGCTGATACACTGAT 80
QY 4150 TTTTATTAACGAAATTAAGATCTAACAGCTAAACAGTAACTGATMAATTTCC 4209
DB 81 TTTTATTAACGAAATTAAGATCTAACAGCTAAACAGTAACTGATMAATTTCC 140
QY 4210 TATACATAGACCCGTTGCTACTCTCAGACCCCTCTCCCAATTTTTTCCGTAGCAT 4269
DB 141 TATACATTAACCCGTTGCTACTCTCAGACCCCTCTCCCAATTTTTTCCGTAGCAT 200
QY 4270 GTGATCCGTGATTAACATCTATTTTCATTGCTTTTCTAATATGGAACAATGAGAG 4329
DB 201 GTGATCCGTGATTAACATCTATTTTCATTGCTTTTCTAATATGGAACAATGAGAG 260
QY 4330 TGAACCTTAATATAGTGTGTAGTATTAACATCTATGACCTAATATTAGAAATGCT 4389
DB 261 TGAACCTTAATATAGTGTGTAGTATTAACATCTATGACCTAATATTAGAAATGCT 320
QY 4390 AATTAGTACGACACATAGAAACATGAATTCCTAGCATGTACCTTTGACACCAT 4449
DB 321 AATTAGTACGACACATAGAAACATGAATTCCTAGCATGTACCTTTGACACCAT 380
QY 4450 TTTG 4453
DB 381 TATG 384
RESULT 13
AAF93519
ID AAF93519 standard; cDNA; 547 BP.
XX
AC AAF93519;

XX
XX 21-MAY-2001 (first entry)
XX
XX cDNA encoding SRT protein isolated from human retina tissue SEQ ID 340.
DE
XX Human; SRT; gene therapy; gene mapping; tissue typing; ss.
XX
XX Homo sapiens.
XX
XX WO200107611-A2.
XX
XX 01-FEB-2001.
XX
XX 21-JUL-2000; 2000WO-US20006.
XX
XX 26-JUL-1999; 99US-0145701.
XX
XX (GEN) GENENTECH INC.
XX
XX Baker KP, Goddard A, Wood WT;
XX WPI; 2001-112729/12.
XX
XX New isolated nucleic acid molecule encoding a SRT polypeptide is useful
PT for production of recombinant SRT polypeptides, gene mapping,
PT diagnosing genetic disorders and for gene therapy -
XX
XX Claim 2; Flg 340; 663pp; English.
XX
XX Sequences AAF93180 - AAF93743 represent polynucleotide sequences encoding
CC human SRT proteins. The cDNA sequences are isolated from various
CC different human tissue cDNA libraries. The invention relates to a method
CC for detecting cDNA encoding an SRT protein, a vector containing cDNA
CC encoding SRT, a host cell transformed with the vector, an isolated SRT
CC polypeptide, and an antibody which binds to SRT. The polynucleotide
CC sequence can be used in gene therapy and is useful in the recombinant
CC production of SRT polypeptides, as a hybridisation probe to screen
CC libraries to isolate cDNAs with sequence identity to SRT polypeptides, to
CC map the gene encoding the SRT polypeptides and analysing genetic
CC disorders, tissue typing and disease tissue detection. The SRT
CC polynucleotide sequences can be used in polymerase chain reaction,
CC screening for new therapeutic molecules and generation of antisense RNA
CC and DNA.
XX
XX Sequence 547 BP; 155 A; 117 C; 131 G; 139 T; 5 other;
SQ
Query Match 6.4%; Score 290; DB 22; Length 547;
Best Local Similarity 88.8%; Pred. No. 2, 3e-57;
Matches 348; Conservative 0; Mismatches 37; Indels 7; Gaps 3;
QY 111 CCTCTCCGTCGCGTGGCGGGGAACTGTGACCGCGCGCTCGGGAACGCCAGGT 170
DB 114 CCTCTCCGTCGCGTGGCGGGGAACTGTGACCGCGCGCTCGGGAACGCCAGGT 172
QY 171 CCCCGCCGAGGTCCCGGGGAGATACATAGATCATCTAGTAAACCTTTGAAGTGG 230
DB 173 CCCCGCCGAGGTCCCGGGGAGATACATAGATCATCTAGTAAACCTTTGAAGTGG 228
QY 231 TTCAGAAATTTTGAATAGCAAAATAGAAATTAAGAAATTAACAGAGATACAGAG 290
DB 229 TTCAGAAATTTTGAATAGCAAAATAGAAATTAAGAAATTAACAGAGATACAGAG 288
QY 291 ACAGATGGAAGTGTCTTAGAAACAGAACAGACAGTGAAGAAAAACAGACAAATCC 350
DB 289 --CAGCATGAAGTGTCTTAGAAACAGAACAGACAGTGAAGAAAAACAGACAAATCC 346
QY 351 GCTCAGATACAACTGACGTGATATGTTTCCGGCTTAATGCTTTAGAGTTGGAGTC 410
DB 347 GCTCAGATACAACTGACGTGATATGTTTCCGGCTTAATGCTTTAGAGTTGGAGTC 406
QY 411 TCTTTTGCATATGTCATTTTATACATGCAACAGTAACTCTTTACAGACAGT 470
DB 407 TCTTTTGCATATGTCATTTTATACATGCAACAGTAACTCTTTACAGACAGT 466

OY 471 CCTCAGAAATATTTAGTACATTGCACAGG 502
|||||
Db 467 CCTCAGAAATATTTAGTACATTGCACAGG 498

RESULT 14
ABV08367
ID ABV08367 standard; cDNA: 447 BP.
XX
AC ABV08367;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 8358.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PE 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI: 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 1328; 11750pp; English.

CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX
SQ Sequence 447 BP; 116 A; 86 C; 75 G; 168 T; 2 other;

Query Match 5.2%; Score 234.6; DB 23; Length 447;
Best Local Similarity 92.6%; Pred. No. 1.6e-44;
Matches 301; Conservative 1; Mismatches 15; Indels 8; Gaps 5;

OY 4089 ATCTACTGATATCTTCTTTAAATTTCTTCACATATGTTGCCGATACAACTGA 4148
|||||
Db 43 ATCTACTGATATCTTCTTTAAATTTCTTCACATATGTTGCCGATACAACTGA 102

OY 4149 TTTTATTAAGTAATTTAAGGAATCTAAGCTAAACCTCAGTAAGTCATMTATTCC 4208

Db 103 TTTTATTAAGTAATTTAAGGAATCTAAGCTAAACCTCAGTAAGTCATMTATTCC 162
|||||

OY 4209 TTATTA-CATGAGACCGTGTGCTACTGTC-AGACACCTCTCCCAATTTTTTCTGTAG 4266
|||||

Db 163 TTATTAACCATAGACCCGCGGTACTCTCAAGCACCTCTCCCAATTTTTTCTGTAG 222

OY 4267 CATGAGACCGTGTGCTACTGTC-AGACACCTCTCCCAATTTTTTCTGTAG 4321
|||||

Db 223 CATGAGACCGTGTGCTACTGTC-AGACACCTCTCCCAATTTTTTCTGTAG 282

OY 4322 AATGAGAGTGAATCTAAATATA-GTTGTAGTAAATTAACATCAATTAATATTA 4380
|||||

Db 283 TTGAGAGTGAATCTAAATATA-GTTGTAGTAAATTAACATCAATTAATATTA 342

OY 4381 GAAATGCTAATTAAGTACACGAC 4405
|||||

Db 343 GAAATGCTAATTAAGTACACGAC 367

RESULT 15
AAS26179
ID AAS26179 standard; cDNA: 1423 BP.
XX
AC AAS26179;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human cDNA encoding a novel secreted protein, Seq ID 358.
XX
KW Human; immunosuppressive; antiarthritic; ss; antihematic;
KW cytostatic; cardiant; vasotropic; cerebroprotective; nocrotic;
KW neuroprotective; antibacterial; virucide; fungicide; optalmalogical;
KW vulnerable; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.

XX
OS Homo sapiens.
XX
PN WO20015322-A2.
XX
PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01341.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226688.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-023142.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235634.
 PR 27-SEP-2000; 2000US-0235636.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239335.
 PR 13-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI: 2001-488783/53.
 DR P-PSDB: AAU16192.
 XX
 PT New nucleic acid molecules encoding 461 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 XX
 PS Claim 1; SEQ ID NO 358; 980pp; English.
 XX
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage

CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence encodes a novel secreted protein of the invention.

Query Match 2.38; Score 106; DB 22; Length 1423;

Best Local Similarity 63.28; Pred. No. 2e-14;

Matches 230; Conservative 0; Mismatches 125; Indels 9; Gaps 4;

```
QY 2844 TAATTTATTTCCCTAAAGCATATATTTAATCATTTCAAGTTTGCAGACTAGTGGCCATCC 2903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 651 TAACTTTGGCTACTACTCAATATATATACCACTTAATATATATGTTCTAGAGCAGCGCTGCC 710
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2904 AATAGAAATTAATAATATAGTCACATATTTTATTAATAATTTCTAGTAACATTAAC 2963
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 711 AATAGAAATTAATAATCTGAGCCACATGTAATTTT--ATTTCTCTAGCCACATTAATA 768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2964 AAAGTAAAGTGACAGCGCAAAA-----TAATTTGATATTACTTTTCACCCAGTAGTA 3018
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 769 AAGTAAAGATACAGACAGACTAATTTTATGTTTAAATTCAGTATATCCAAAATATC 828
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3019 TACCCAAAATAGCGAAATATATAGAAATTTATATGAGATATTTTACATCCTTTT-TTGTAC 3077
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 829 ATTGAACATGTATATATATATAAATTTATATGATATTTTACATCTTTTGGTAATA 888
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3078 CAAGTCTTTTAATAGCAGTACATATTTTATCTTACTGCAATTTCTTACTCCGAGTAGCC 3137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 889 CTAGTCTTCAAAATCTGATGATGTATCTACATGTGATAGCACATCTCAGCTTGTACTAGCC 948
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3138 ATATTTCAAGTGTCAT-TTGGCACATGCGGCTGAGTACTGATTTGGACAGTTCAGTA 3196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 949 ACATTTGCAAGTGTGAGTACACATGTGGCTAGTGGCTACTGACATGACAGCACAGTT 1008
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3197 CTAG 3200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1009 CTAG 1012
```

Search completed: March 22, 2003, 08:54:57
Job time : 1003 secs

THIS PAGE BLANK (USPTO)

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2003, 03:03:41 ; Search time 12776 Seconds

(without alignments)
10309.901 Million cell updates/sec

Title: US-09-847-046-1

Sequence: 1 gccgagtcggtgcgctgc.....cctcaaaaaaaaaa 4526

Scoring table: IDENTITY_NUC

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 41.09280

| | |
|------------------------|---|
| Minimum DB seq length: | 0 |
|------------------------|---|

Post-processing: Minimum Match 08

Maximum Match 100%
Listing first 45 summaries

Database

```

1:  gb_ba:*
2:  gb_hlg:*
3:  gb_in:*
4:  gb_om:*
5:  gb_ov:*
6:  gb_pat:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: gb_vl:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_in:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pln:*
35: em_hlg_rod:*
36: em_hlg_nam:*
37: em_hlg_vrt:*
38: em_sy:*
39: em_hgo_hum:*
40: em_hgo_mus:*
41: em_hgo_other:*

```

Pred. No. is the number of results predicted by chance to have a

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|--------|-------------|--------|----|----------|--------------------|
| 1 | 4433.6 | 98.0 | 4552 | | AB037765 | AB037765 Homo sapi |
| 2 | 2811.6 | 62.1 | 2820 | | AK022563 | AK022563 Homo sapi |
| 3 | 2410.2 | 55.3 | 2747 | | AK097748 | AK097748 Homo sapi |
| 4 | 1995.8 | 44.1 | 16100 | 2 | AC028659 | AC028659 Homo sapi |
| 5 | 1995.8 | 44.1 | 176368 | 1 | BC021RGC | AL157971 Human chr |
| 6 | 426.4 | 9.4 | 1496 | 10 | BC027108 | BC027108 Mus muscu |
| 7 | 307.2 | 6.8 | 204606 | 2 | AC021131 | AC021131 Homo sapi |
| 8 | 290 | 6.4 | 547 | 6 | AX079596 | AX079596 Sequence |
| 9 | 233.6 | 5.2 | 204606 | 2 | AC021131 | AC021131 Homo sapi |
| 10 | 201.4 | 4.4 | 212103 | 2 | CNS01DV4 | AL133453 Human chr |
| 11 | 129.4 | 2.9 | 203610 | 2 | AC102579 | AC102579 Mus muscu |
| 12 | 126.6 | 2.8 | 203610 | 2 | AC102579 | AC102579 Mus muscu |
| 13 | 124 | 2.7 | 94749 | 2 | HS050N11 | AL031122 Human DNA |
| 14 | 115 | 2.5 | 14499 | 2 | AC094089 | AC094089 Homo sapi |
| 15 | 111 | 2.5 | 89196 | 2 | AC012619 | AC012619 Homo sapi |
| 16 | 111 | 2.5 | 100457 | 2 | AC020933 | AC020933 Homo sapi |
| 17 | 111 | 2.5 | 132641 | 2 | AC012605 | AC012605 Homo sapi |
| 18 | 111 | 2.5 | 154054 | 2 | AC020919 | AC020919 Homo sapi |
| 19 | 111 | 2.5 | 155766 | 2 | AC114961 | AC114961 Homo sapi |
| 20 | 111 | 2.5 | 170970 | 2 | AC021564 | AC021564 Homo sapi |
| 21 | 111 | 2.5 | 264095 | 2 | AC008500 | AC008500 Homo sapi |
| 22 | 110.2 | 2.4 | 172388 | 2 | AC027634 | AC027634 Homo sapi |
| 23 | 110.2 | 2.4 | 176814 | 2 | AC100849 | AC100849 Homo sapi |
| 24 | 109.2 | 2.4 | 174710 | 2 | AC069134 | AC069134 Homo sapi |
| 25 | 109.2 | 2.4 | 176807 | 2 | AC037487 | AC037487 Homo sapi |
| 26 | 108.4 | 2.4 | 35714 | 9 | HS198BP4 | AL008708 Human DNA |
| 27 | 107.2 | 2.4 | 158305 | 9 | CNS07EFA | AL151261 Human chr |
| 28 | 106 | 2.3 | 173043 | 9 | AC068794 | AC068794 Homo sapi |
| 29 | 106 | 2.3 | 196710 | 2 | AC126470 | AC126470 Homo sapi |
| 30 | 105.6 | 2.3 | 165845 | 2 | AC019054 | AC019054 Homo sapi |
| 31 | 105.6 | 2.3 | 171947 | 9 | AC091045 | AC091045 Homo sapi |
| 32 | 105.6 | 2.3 | 192100 | 9 | AC021810 | AC021810 Homo sapi |
| 33 | 105.4 | 2.3 | 140356 | 9 | AC007628 | AC007628 Genomic S |
| 34 | 105.4 | 2.3 | 200829 | 2 | AC090208 | AC090208 Homo sapi |
| 35 | 105.2 | 2.3 | 158690 | 2 | AC130197 | AC130197 Fells cat |
| 36 | 105 | 2.3 | 118985 | 9 | AC094102 | AC094102 Homo sapi |
| 37 | 105 | 2.3 | 176552 | 2 | AC024622 | AC024622 Homo sapi |
| 38 | 105 | 2.3 | 178530 | 2 | AC116362 | AC116362 Homo sapi |
| 39 | 104.8 | 2.3 | 146258 | 2 | AC023479 | AC023479 Homo sapi |
| 40 | 104.8 | 2.3 | 160012 | 2 | AL136231 | AL136231 Human DNA |
| 41 | 104.6 | 2.3 | 167886 | 2 | AP002405 | AP002405 Homo sapi |
| 42 | 104.6 | 2.3 | 170321 | 9 | AC097466 | AC097466 Homo sapi |
| 43 | 104.4 | 2.3 | 184332 | 2 | AC011777 | AC011777 Homo sapi |
| 44 | 103.4 | 2.3 | 200610 | 9 | AF002851 | AF002851 Homo sapi |
| 45 | 103.8 | 2.3 | 122674 | 9 | AC007021 | AC007021 Homo sapi |

ALIGNMENTS

| | | | | | | |
|------------|---|----------|-------------|------|--------|-----------------|
| RESULT 1 | AB037765 | AB037765 | 4552 bp | mRNA | linear | PRI 10-MAY-2002 |
| LOCUS | AB037765 | | | | | |
| DEFINITION | Homo sapiens mRNA for KIAA1344 protein, partial cds. | | | | | |
| ACCESSION | AB037765 | | | | | |
| VERSION | AB037765.2 | | GI:20521893 | | | |
| KEYWORDS | | | | | | |
| SOURCE | Homo sapiens brain cDNA to mRNA, clone_lib:pbJuncscriptII SK plus clone:rfj00476s1. | | | | | |
| ORGANISM | Homo sapiens | | | | | |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | | |
| REFERENCE | 1 Nagase,T., kikuno,R., Ishikawa,K.I., Hirosewa,M. and Ohara,O. Prediction of the coding sequences of unidentified human genes. | | | | | |
| AUTHORS | | | | | | |
| TITLE | | | | | | |

[illegible][illegible]

| | | | |
|----------|----------|---|-----------------------------|
| OY | 3500 | TGCTGTCGTGGCAGGCGTGAATACCTGTGTAAAGTAACCTGTGGCAACATGATT | 3559 |
| Db | 3541 | TGGTGTGTCGTGCCAAGGCTGAGTAACCTGTGTAAAGTAACCTGTGGCAACATGATT | 3600 |
| OY | 3560 | TCCAAATTTCTGGTTCGTGCTCAGGTATCATGTGTGAAAATTTTGGCTATTAAAGAT | 3619 |
| Db | 3601 | TCCAAATTTCTGGTTCGTGCTCAGGTATCATGTGTGAAAATTTTGGCTATTAAAGAT | 3660 |
| OY | 3620 | ATGTATTAGATGGTCTTATCTGATTATTACTGGATACACTTGATCTTTTCTAATATT | 3679 |
| Db | 3661 | ATGTATTAGATGGTCTTATCTGATTATTACTGGATACACTTGATCTTTTCTAATATT | 3720 |
| OY | 3680 | TTCCAGAAATGATNGGATTTAACCTTGAAGAAGACCTCAGAAATATTTATTATTTAAAGT | 3739 |
| Db | 3721 | TTCCAGAAATGATNGGATTTAACCTTGAAGAAGACCTCAGAAATATTTATTATTTAAAGT | 3780 |
| OY | 3740 | AGCTTTAAAAACCTCTCTTATTATTCTACAAGTTATATAGCTAAATTTTCAGATTGAACAGGG | 3799 |
| Db | 3781 | AGCTTTAAAAACCTCTCTTATTATTCTACAAGTTATATAGCTAAATTTTCAGATTGAACAGGG | 3840 |
| OY | 3800 | ATTCCAGCATTTGCCATCTCCCTCATGGAAGAAGAGGCTCCCATCTGAAAGGCTCTCTGA | 3859 |
| Db | 3841 | ATTCCAGCATTTGCCATCTCCCTCATGGAAGAAGAGGCTCCCATCTGAAAGGCTCTCTGA | 3900 |
| OY | 3860 | AATCTACCCCTTCAAGCTTCACAGCAAAATCAGTTGATCTCCCTGAGCACAACGGCTCATTT | 3919 |
| Db | 3901 | AATCTACCCCTTCAAGCTTCACAGCAAAATCAGTTGATCTCCCTGAGCACAACGGCTCATTT | 3960 |
| OY | 3920 | CTGTGAGGAGGAGGAAGATTAGCCAAAGAGTTAATTTTTCATTCCAAATCATTAGCTGTT | 3979 |
| Db | 3961 | CTGTGAGGAGGAGGAAGATTAGCCAAAGAGTTAATTTTTCATTCCAAATCATTAGCTGTT | 4020 |
| OY | 3980 | AGACGATCTGTTTGTAGAGATGTTGTGCTCATTTTTCCTCGAGCATTTTGTGGACA | 4039 |
| Db | 4021 | AGACGATCTGTTTGTAGAGATGTTGTGCTCATTTTTCCTCGAGCATTTTGTGGACA | 4080 |
| OY | 4040 | TTTGTGGAAATATTCTAATTTGGTCTACTGTATTTTTCTTTTAAATATCTACTGAT | 4099 |
| Db | 4081 | TTTGTGGAAATATTCTAATTTGGTCTACTGTATTTTTCTTTTAAATATCTACTGAT | 4140 |
| OY | 4100 | ATCTTGTTCTTTAAATTTTCTTCACATATGAGTTTGCCCTGATACCACTGATTTTATACT | 4159 |
| Db | 4141 | ATCTTGTTCTTTAAATTTTCTTCACATATGAGTTTGCCCTGATACCACTGATTTTATACT | 4200 |
| OY | 4160 | GAATATTTAAGAACTTAACAGCTTAACAACTCAGTAAGTGATATTTTCTCTTAATACATAG | 4219 |
| Db | 4201 | GAATATTTAAGAACTTAACAGCTTAACAACTCAGTAAGTGATATTTTCTCTTAATACATAG | 4260 |
| OY | 4220 | ACCCTGTGCTACTCTCAGACACCCTCTCTCAATTTTTTTTCTGTAGCATGTGATGCTGT | 4279 |
| Db | 4261 | ACCCTGTGCTACTCTCAGACACCCTCTCTCAATTTTTTTTCTGTAGCATGTGATGCTGT | 4320 |
| OY | 4280 | ATTAAACATCATTTTCAATTTTGTCTTTATTTCTAATATGGAACAATAGAGGTAACTCTTA | 4339 |
| Db | 4321 | ATTAAACATCATTTTCAATTTTGTCTTTATTTCTAATATGGAACAATAGAGGTAACTCTTA | 4380 |
| OY | 4340 | ATATAGGTTGTGATTAATAAATCATCTTAGCTAATATTATTAGAAAATGCTAATTAAGTAC | 4399 |
| Db | 4381 | ATATAGGTTGTGATTAATAAATCATCTTAGCTAATATTATTAGAAAATGCTAATTAAGTAC | 4440 |
| OY | 4400 | CAGCACAATAGAAAATGAAATTTGCTTAGTCATTTGACCTTTGTGAGCAATTTTGAACAGTC | 4459 |
| Db | 4441 | CAGCACAATAGAAAATGAAATTTGCTTAGTCATTTGACCTTTGTGAGCAATTTTGAACAGTC | 4500 |
| OY | 4460 | ATTAAATGTTTGTCTAATATTTTAAATAAAGTGTCTGGTTTTCGAATACCTTC | 4511 |
| Db | 4501 | ATTAAATGTTTGTCTAATATTTTAAATAAAGTGTCTGGTTTTCGAATACCTTC | 4552 |
| RESULT 2 | | | |
| LOCUS | AK022563 | 2820 bp | mRNA linear PRI 01-AUG-2002 |

| | DEFINITION | Homo sapiens CDNA FLJ12501 f1s, clone NT2RM2001681. |
|----------------------------|---|---|
| ACCESSION | AK022563 | |
| VERSION | AK022563.1 | GI:10434024 |
| KEYWORDS | cdna capping; f1s (full insert sequence); | |
| SOURCE | Homo sapiens testicular carcinoma cell_line:NT2 | CDNA to mRNA, clone_1lib:NT2RM2 clone:NT2RM2001681. |
| ORGANISM | Homo sapiens | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | |
| AUTHORS | Isoqai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y., Matsumura,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahori,K., Masuno,Y. and Sasaki,N. NEO human cDNA sequencing project Unpublished 2 (bases 1 to 2820) Isoqai,T. and Otsuki,T. Direct Submission Submitted (23-AUG-2000) Takao Isoqai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 297-0812, Japan (E-mail:genom@helix.co.jp, Tel:01-438-52-3975, Fax:01-438-52-3986) | |
| JOURNAL | NEO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; CDNA full insert sequencing: Research Association for Biotechnology; CDNA library construction, 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo. | |
| COMMENT | Location/Qualifiers 1..2820 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="NT2RM2001681" /cell_line="NT2" /cell_type="teratocarcinoma" /clone_lib="NT2RM2" /note="cloning vector: pME18SFL3-mRNA from uninduced NT2 neutonal precursor cells." 40..1113 /note="unnamed protein product" /codon_start=1 /protein_id="PAB14101.1" /db_xref="GI:10434025" /translation="MYKKGEPPVSYAGMLGTEDELKFTQLNRISYPVNITSGAEEI LSCGLYKDLILYSSVSVLGPSPTMKRAEDFSAGNLKGIVTGYISEDDVLLST KYAAALPALRLARTBEKIEBIPLASHTADIDVOITDALLEMPETVEENLPSEFRI OKPLIFESDGTVPQPKAILTLVKRKYLDSPCWLNLKTPGVGRILAYAPDPLE PIPLSLVINHSGGVAFAPPSDOAIIEENVLMKTI.EAGLENIHTTILPADEMRPELE AYPLELSMDATDSORCTRYKRVKPKCKKEFDVQENDEKHEDSAVKEPTEFLRTIKHMNR SNMFKEAKSFRRDKELGCSKV" | |
| CDS | | |
| BASE COUNT | 911 a 496 c 504 g 909 t | |
| ORIGIN | | |
| Query Match | 62.1%; Score 2811.6; | DB 9; Length 2820; |
| Best Local Similarity | 99.8%; Pred. No. 0; | |
| Matches 2814; Conservative | 1; Mismatches | 5; Indels 0; Gaps 0; |
| Db | 1 | TGTTACTAAGCAAAATGTTACTGAATTCCTATCATATAAAGATGTACACAAGAAGCGAGAAC 1745 |
| Qy | 1686 | TGTACTAAGCAAAATGTTACTGAATTCCTATCATATAAAGATGTACACAAGAAGCGAGAAC 1745 |
| Dd | 1 | TGTACTAAGCAAAATGTTACTGAATTCCTATCATATAAAGATGTACACAAGAAGCGAGAAC 60 |
| Qy | 1746 | CCAGTATTCTTAGTGCGAATGTTAGAACCAAAGATCTCTAAAATTTATTCACAGCTCAAC 1805 |
| Dd | 61 | CCAGTATTCTTAGTGCGAATGTTAGAACCGAAGATCTCTAAAATTTATTCACAGCTCAAC 120 |
| Qy | 1806 | AGGATTTCATATCCAGTGAATATTAATCAGATCCAGAAGCAAGAAGATATTTAAGTGGG 1865 |
| Dd | 121 | AGGATTTCATATCCAGTGAATATTAATCAGATCCAGAAGCAAGAAGATATTTAAGTGGG 180 |
| Qy | 1866 | GAAATTATTAANAAGCCTCATCTTGATATCTAGTGTGCAGATTTGGGACTATTAGTCCA 1925 |
| Dd | 181 | GAAATTATTAANAAGCCTCATCTTGATATCTAGTGTGCAGATTTGGGACTATTAGTCCA 240 |

| | | | |
|----|------|--|------|
| QY | 1926 | ACCATGAAAAACAGCAAAAGACATTTTGTAGAACAGAGAAACTACTTAAAGATATGTT | 1985 |
| Db | 241 | ACCATGAAAAACAGCAAAAGACATTTTGTAGAACAGAGAAACTACTTAAAGATATGTT | 300 |
| QY | 1986 | ATCAGCTGGAATTTATTTCTGAGAGAAGATGTTTGTCTACTGTCAACCAATATGTCTGCAGT | 2045 |
| Db | 301 | ATCAGCTGGAATTTATTTCTGAGAGAAGATGTTTGTCTACTGTCAACCAATATGTCTGCAGT | 360 |
| QY | 2046 | CTTCCAGCCCTGCTGCTGTCGCAACACACAGAAAGGCAAAATAGAGACATCCCACTAGCT | 2105 |
| Db | 361 | CTTCCAGCCCTGCTGCTGTCGCAACACACAGAAAGGCAAAATAGAGACATCCCACTAGCT | 420 |
| QY | 2106 | AGCACACATGCACACAGACATAGTTCAATTAATAACAGATCCAGTACTAGGAATGTTTCCG | 2165 |
| Db | 421 | AGCACACATGCACACAGACATAGTTCAATTAATAACAGATCCAGTACTAGGAATGTTTCCG | 480 |
| QY | 2166 | GAATTCAGCTGTGGAATAATCTTCCAGTATTTTCAGACTTCAGAAACCATTAATGATTTTG | 2225 |
| Db | 481 | GAATTCAGCTGTGGAATAATCTTCCAGTATTTTCAGACTTCAGAAACCATTAATGATTTTG | 540 |
| QY | 2226 | TTTCAGTATGGCAGCTGATAAATCCTCAATATTAATAAAGCATATTGACACAGGTAAACAG | 2285 |
| Db | 541 | TTTCAGTATGGCAGCTGATAAATCCTCAATATTAATAAAGCATATTGACACAGGTAAACAG | 600 |
| QY | 2286 | AAATACCTTGGAATTCATTACTCCATGCTGGTTAAATCTTAAAGATACTCCAGTGGGAGA | 2345 |
| Db | 601 | AAATACCTTGGAATTCATTACTCCATGCTGGTTAAATCTTAAAGATACTCCAGTGGGAGA | 660 |
| QY | 2346 | GGAACTTTGCGGGCATATTTTGTATCTCTGTGCTCCCTTCTCTTCTTGTGTTGGTGAAT | 2405 |
| Db | 661 | GGAACTTTGAGGGCATATTTTGTATCTCTGCTCCCTCCCTTCTCTTGTGTTGGTGAAT | 720 |
| QY | 2406 | CTGCATTTACAGGTGGCCACAGATATTGATTTCCCTTCAGACACAGGCTAATATGGAAGAAAC | 2465 |
| Db | 721 | CTGCATTTACAGGTGGCCACAGATATTGATTTCCCTTCAGACACAGGCTAATATGGAAGAAAC | 780 |
| QY | 2466 | CTTTGATTTGTGGCTGGAATAATTAGAAGCAGACTAGAAAAATCATATCACAAATTTTACCT | 2525 |
| Db | 781 | CTTTGATTTGTGGCTGGAATAATTAGAAGCAGACTAGAAAAATCATATCACAAATTTTACCT | 840 |
| QY | 2526 | GCTCAAGAATGGAACCTCTCTCTCCAGCTTATGATTTTCTAAGTATGATAGATGCCGCA | 2585 |
| Db | 841 | GCTCAAGAATGGAACCTCTCTCTCCAGCTTATGATTTTCTAAGTATGATAGATGCCGCA | 900 |
| QY | 2586 | ACATTCACAAGTGGCAGCTGGAAGAAGTCCCAACTGTATGAAAGAAACAGATGTGCAGAG | 2645 |
| Db | 901 | ACATTCACAAGTGGCAGCTGGAAGAAGTCCCAACTGTATGAAAGAAACAGATGTGCAGAG | 960 |
| QY | 2646 | AATGATTAAGAACACACATAGATTAATGTGGCAGTGAAGAAACCCGATTGGAACCTGTG | 2705 |
| Db | 961 | AATGATTAAGAACACACATAGATTAATGTGGCAGTGAAGAAACCCGATTGGAACCTGTG | 1020 |
| QY | 2706 | AGAAATTAAGCAATTTGGAATTAAGAGTAATTTGGTTTAAAGAGCAGAAAAATCATTTAGACGT | 2765 |
| Db | 1021 | AGAAATTAAGCAATTTGGAATTAAGAGTAATTTGGTTTAAAGAGCAGAAAAATCATTTAGACGT | 1080 |
| QY | 2766 | GATTAAGAGATTTGGAGATGCTCAAAAGAGAACTAATTTTATGAGGCTGTGGTTCCAAAATT | 2825 |
| Db | 1081 | GATTAAGAGATTTGGAGATGCTCAAAAGAGAACTAATTTTATGAGGCTGTGGTTCCAAAATT | 1140 |
| QY | 2826 | TTTTTGGCATGATAGACTTAATTTATTTCTCTTAAGCAATTAATTAATCATTTTCAAGTT | 2885 |
| Db | 1141 | TTTTTGGCATGATAGACTTAATTTATTTCTCTTAAGCAATTAATTAATCATTTTCAAGTT | 1200 |
| QY | 2886 | TGCAGACTGTGTCATCCAAATAGAAATTAATTAATTAAGTCACAATTTTATTTAAATTTT | 2945 |
| Db | 1201 | TGCAGACTGTGTCATCCAAATAGAAATTAATTAATTAAGTCACAATTTTATTTAAATTTT | 1260 |
| QY | 2946 | CTAGTATACATCAATTAAACAAAGTAAGTAGAGAGGCAAAATTAATTTGATATTTACTTT | 3005 |
| Db | 1261 | CTAGTATACATCAATTAAACAAAGTAAGTAGAGAGGCAAAATTAATTTGATATTTACTTTT | 1320 |

| | | | |
|----|------|--|------|
| QY | 3006 | TCGCCAGTACTATATCCCAAAATATAGCAAAATATAGAAATTTATTAATGAGATATTTTACAT | 3065 |
| Db | 1321 | TCACCAGTACTATATCCCAAAATATAGCAAAATATAGAAATTTATTAATGAGATATTTTACAT | 1360 |
| QY | 3066 | CGTTTTTGTATGCCAAGCTCTCTAAATGACATATATTTATATAGTACATGCAATTTCTTAC | 3125 |
| Db | 1381 | CGTTTTTGTATGCCAAGCTCTCTAAATGACATATATTTATATAGTACATGCAATTTCTTAC | 1440 |
| QY | 3126 | TTCCGAGTAGGCATATTTTCAAGTGTTCATTTGGCCACATGTGGCCTGTGACTACTGATATGG | 3185 |
| Db | 1441 | TTCCGAGTAGGCATATTTTCAAGTGTTCCTTTGCCACATGTGGCCTGTGACTACTGATATGG | 1500 |
| QY | 3186 | ACAGTTCAGTACTAGACAAAACATAGCATATTTAACTTATGTTCTAGCCATGATTTCTATTT | 3245 |
| Db | 1501 | ACAGTTCAGTACTAGACAAAACATAGCATATTTAACTTATGTTCTAGCCATGATTTCTATTT | 1560 |
| QY | 3246 | TGGATTTAAATTTAACTCTAATCAGATTTACCTCCACATGATTCATGCACTGACAGT | 3305 |
| Db | 1561 | TGGATTTAAATTTAACTCTAATCAGATTTAACTCCACATGATTCATGCACTGACAGT | 1620 |
| QY | 3306 | TATATTTGTTTTTATTTGGAGTCATGATATTTAAATTCAGCCTTTGTCAACCTCAGGGAGAT | 3365 |
| Db | 1621 | TATATTTGTTTTTATTTGGAGTCATGATATTTAAATTCAGCCTTTGTCAACCTCAGGGAGAT | 1680 |
| QY | 3366 | TTAGCAATTTGTGGGAGACATTTTTTGATGTCAATGACTAGGGCAGTTATTTAGCATTTAGT | 3425 |
| Db | 1681 | TTAGCAATTTGTGGGAGACATTTTTTGATGTCAATGACTAGGGCAGTTATTTAGCATTTAGT | 1740 |
| QY | 3426 | AGTAGAGGCCAATGGATTCCTCTAAATTAACCTGATTTGGACAGCCGCCACAAACAAAGAT | 3485 |
| Db | 1741 | AGTAGAGGCCAATGGATTCCTCTAAATTAACCTGATTTGGACAGCCGCCACAAACAAAGAT | 1800 |
| QY | 3486 | TATCTCGCCCAAAATGGTATGTCGTGCCAAGGCGAGATACCTTGTTTAAAAGTAACTGT | 3545 |
| Db | 1801 | TATCTCGCCCAAAATGGTATGTCGTGCCAAGGCGAGATACCTTGTTTAAAAGTAACTGT | 1860 |
| QY | 3546 | TGGCAGACTAGCTTCCAGAAATTTCTGTCTGTCTCACGTATCATGTTTGAAAAAATTT | 3605 |
| Db | 1861 | TGGCAGACTAGCTTCCAGAAATTTCTGTCTGTCTCACGTATCATGTTTGAAAAAATTT | 1920 |
| QY | 3606 | TGGCATTTAAAGATATGATTTAGATGGTCTTATACCGATTTATTTACCTGGATTCACACTTGA | 3665 |
| Db | 1921 | TGGCATTTAAAGATATGATTTAGATGGTCTTATACCGATTTATTTACCTGGATTCACACTTGA | 1980 |
| QY | 3666 | TCTTTTCTAATATTTTACAGAAATGATGAGGATTAACCTTAGAGAGAGACTCAGAAATGATAT | 3725 |
| Db | 1981 | TCTTTTCTAATATTTTACAGAAATGATGAGGATTAACCTTAGAGAGAGACTCAGAAATGATAT | 2040 |
| QY | 3726 | TTATATTTTAAAGTAGAGCTTAAACACCTCTCTATTTTCTACAAAGTTATATGCTAAATTT | 3785 |
| Db | 2041 | TTATATTTTAAAGTAGAGCTTAAACACCTCTCTATTTTCTACAAAGTTATATGCTAAATTT | 2100 |
| QY | 3786 | CAGATTTGAACAGGGATTCAGACATTTCTGCGCATCTCTCATGGAAGAGAGGCTCCCTCATC | 3845 |
| Db | 2101 | CAGATTTGAACAGGGATTCAGACATTTCTGCGCATCTCTCATGGAAGAGAGGCTCCCTCATC | 2160 |
| QY | 3846 | TGAAGCGCTGTGAAATCTACCTTCGCAACCTTCAGACAATACAGTATCTCCCTGAGC | 3905 |
| Db | 2161 | TGAAGCGCTGTGAAATCTACCTTCGCAACCTTCAGACAATACAGTATCTCCCTGAGC | 2220 |
| QY | 3906 | CACACGGCCTCATTTCTGTGAGGAGGGAAGATTTAGCCAAAGAGTTAATTTTCATTTCCAA | 3965 |
| Db | 2221 | CACACGGCCTCATTTCTGTGAGGAGGGAAGATTTAGCCAAAGAGTTAATTTTCATTTCCAA | 2280 |
| QY | 3966 | ATCACCTTAGCTTTAGACCTCATCTGTTTGTAGACAGTTGTTTGTCTCATTTTTTGGCTCTGTG | 4025 |
| Db | 2281 | ATCACCTTAGCTTTAGACCTCATCTGTTTGTAGACAGTTGTTTGTCTCATTTTTTGGCTCTGTG | 2340 |
| QY | 4026 | CATTTTTTGAACATTTGTTTGAGAAATATTCCTATTTTGGTCTCTACCTGATATTTTCTTTTT | 4085 |
| Db | 2341 | CATTTTTTGAACATTTGTTTGAGAAATATTCCTATTTTGGTCTCTACCTGATATTTTCTTTTT | 2400 |
| QY | 4086 | AATATCTACTGTGATATCTGTCTTTCTTAAATTTCTTACACATATGGTTTGGCTGATACAC | 4145 |

```

|||||
Db 2401 AATATCTACTTGATATCTTCTTTAAATTTCTTCACATAGTGTGGCTGATACAC 2460
QY 4146 TGATTTTATTAAGTAATTAAGAACTACAGCTAAACCTAGTAAGTGCATMTATT 4205
Db 2461 TGATTTTATTAAGTAATTAAGAACTACAGCTAAACCTAGTAAGTGCATMTATT 2520
QY 4206 TCGTTATTAAGTAATTAAGAACTACAGCTAAACCTAGTAAGTGCATMTATT 4265
Db 2521 TCGTTATTAAGTAATTAAGAACTACAGCTAAACCTAGTAAGTGCATMTATT 2580
QY 4266 GCATGATGATCCGATTAATTAAGTAATTAAGAACTACAGCTAAACCTAGTAAGTGCATMTATT 4325
Db 2581 GCATGATGATCCGATTAATTAAGTAATTAAGAACTACAGCTAAACCTAGTAAGTGCATMTATT 2640
QY 4326 AGATGATGATCCGATTAATTAAGTAATTAAGAACTACAGCTAAACCTAGTAAGTGCATMTATT 4385
Db 2641 AGATGATGATCCGATTAATTAAGTAATTAAGAACTACAGCTAAACCTAGTAAGTGCATMTATT 2700
QY 4386 TCGTTATTAAGTAATTAAGAACTACAGCTAAACCTAGTAAGTGCATMTATT 4445
Db 2701 TCGTTATTAAGTAATTAAGAACTACAGCTAAACCTAGTAAGTGCATMTATT 2760
QY 4446 CAATTTTGACAGTCAATTAAGTGTTCATTAATTTTAAATTAAGTGTTCATTAATTTTAAAT 4505
Db 2761 CAATTTTGACAGTCAATTAAGTGTTCATTAATTTTAAATTAAGTGTTCATTAATTTTAAAT 2820

RESULT 3
AK097748 2747 bp mRNA linear PRI 15-JUL-2002
LOCUS AK097748
DEFINITION Homo sapiens cDNA FLJ40429 f1s, clone TEST12039177.
ACCESSION AK097748
VERSION AK097748.1 GI:21757613
KEYWORDS oligo cloning; f1s (full insert sequence).
SOURCE Homo sapiens testis cDNA to mRNA, clone_11b:TEST12
Clone:TEST12039177.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1
REFERENCE
AUTHORS Oshima, A., Takahashi, Fujii, A., Tanase, T., Inoue, N., Takeuchi, K.,
Arita, M., Musashino, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R.,
Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,
Kawai-Hiro, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,
Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,
Wagatsuma, M., Muraoka, K., Kanehori, K., Sugiyama, A., Kawakami, B.,
Suzuki, Y., Sugano, S., Nagahari, K., Masuno, Y., Nagai, K. and
Isogai, T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2747)
AUTHORS Isogai, T. and Yamamoto, J.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan: cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
Source
Location/Qualifiers
1..2747
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TEST12039177"
/tissue_type="testis"
/clone_11b="TEST12"
/note="cloning vector: PME1BSFL3"

```

```

BASE COUNT 899 a 510 c 575 g 763 t
ORIGIN
Query Match 53.3%; Score 2410.2; DB 9; Length 2747;
Best Local Similarity 94.2%; Pred. No. 0;
Matches 2588; Conservative 0; Mismatches 3; Indels 156; Gaps 2;

QY 139 GTTGGCCGCGCGCTTCGGAGACGGCCAGGTCGCCGCGGAGGTCGCCGAGATAC 198
Db 1 GTTGGCCGCGCGCTTCGGAGACGGCCAGGTCGCCGCGGAGGTCGCCGAGATAC 60
QY 199 ATGATCATCATGTAAGAACTCTTGTAGTGTTCAGAAATTTGAAGTGAAGTGAAGT 258
Db 61 ATGATCATCATGTAAGAACTCTTGTAGTGTTCAGAAATTTGAAGTGAAGTGAAGT 120
QY 259 AGAAATTAAGAACTTAACAGACATACAGAGACAGATGGAAGTGTCTTGAAGAAC 318
Db 121 AGAAATTAAGAACTTAACAGACATACAGAGACAGATGGAAGTGTCTTGAAGAAC 180
QY 319 AGAACACAGCAGTGAAGAAACAGACAAATCCGCTCAGATACAGTGCAGCTGATATGT 378
Db 181 AGAACACAGCAGTGAAGAAACAGACAAATCCGCTCAGATACAGTGCAGCTGATATGT 240
QY 379 TTTCCGGCTTCATGCTTTAGAGTTGGATCTCTTTGCAATATGTCATTTTATTA 438
Db 241 TTTCCGGCTTCATGCTTTAGAGTTGGATCTCTTTGCAATATGTCATTTTATTA 300
QY 439 TGCACACAGTAACTCTTTACAGAACTGCTCAGAAATTTTATTAAGTTCGAC 498
Db 301 TGCACACAGTAACTCTTTACAGAACTGCTCAGAAATTTTATTAAGTTCGAC 360
QY 499 CAGG-----TC 504
Db 361 CAGGAAAGCGCTCTTATAGTTATTTTTCAGACGTATTCGCCAAGACATCTGATTT 420
QY 505 TTGAAGACCTGATGAGGCTGTAGACCTCTGAGAGACATATGAATTTAGTTGCCAAG 564
Db 421 TTGAAGACCTGATGAGGCTGTAGACCTCTGAGAGACATATGAATTTAGTTGCCAAG 480
QY 565 TTATTTGTGCAAGAAAGAAATATCAAGTACTGTGAAAGAAAGATTTGATGAAG 624
Db 481 TTATTTGTGCAAGAAAGAAATATCAAGTACTGTGAAAGAAAGATTTGATGAAG 540
QY 625 CATATTTATTAAGGCAACATATTCCTCAGAGATTTCCCTACTGACACCTGTTGATG 684
Db 541 CATATTTATTAAGGCAACATATTCCTCAGAGATTTCCCTACTGACACCTGTTGATG 600
QY 685 TGAATGCCATTTGCGCCCATGTTCTCTTGTCTCTTTTACTGAAGTGAATATATTA 744
Db 601 TGAATGCCATTTGCGCCCATGTTCTCTTGTCTCTTTTACTGAAGTGAATATATTA 660
QY 745 CCAACCTGGAAGACCTTCGAAACATAGAAATATCTCTGAAAGAAAGCAATATATAT 804
Db 661 CCAACCTGGAAGACCTTCGAAACATAGAAATATCTCTGAAAGAAAGCAATATATAT 720
QY 805 TCTCATATGTAAGAGCATTGGAATACAGAGACAGACAGTGCATGGAAGCGGTTTGG 864
Db 721 TCTCATATGTAAGAGCATTGGAATACAGAGACAGACAGTGCATGGAAGCGGTTTGG 780
QY 865 TGTATGGGACTACATACCAATTTGTCTTAACCAACAAATTTGCCCTTTTGGAAAGTAT 924
Db 781 TGTATGGGACTACATACCAATTTGTCTTAACCAACAAATTTGCCCTTTTGGAAAGTAT 840
QY 925 GCTCTAGAGTGTGAATATGACATCTACTTTTATGTAAGTGAAGTGTGAGT 984
Db 841 GCTCTAGAGTGTGAATATGACATCTACTTTTATGTAAGTGAAGTGTGAGT 900
QY 985 TGACCCAGCAATGTAGAAAGCACTAATAGCAAGCAGCATTGACTGACATTCAC 1044
Db 901 TGACCCAGCAATGTAGAAAGCACTAATAGCAAGCAGCATTGACTGACATTCAC 960
QY 1045 TGTATTATTAAGACATGAAGCACTCTGTGTGACTGAAGTGTGCTGAAGTCTCAACAG 1104
Db 961 TGTATTATTAAGACATGAAGCACTCTGTGTGACTGAAGTGTGCTGAAGTCTCAACAG 920

```

Db 961 TGTATTATAGACAATGAAGCACCCTGTGTACTGAATGCTGTAGATCTTCAACAAG 1020
 Qy 1105 TTTCACGTCTCATCTCCACACTGGGCTTACCACTGGTTTATTTGTATACCACAGGCTA 1164
 Db 1021 TTTCACGTCTCATCTCCACACTGGGCTTACCACTGGTTTATTTGTATACCACAGGCTA 1080
 Qy 1165 CTTATGAAGCTGATGAAGAACTGAGAAATGGGTTGCTTGGGCTCTTGGGAAAAAGCAG 1224
 Db 1081 CTTATGAAGCTGATGAAGAACTGAGAAATGGGTTGCTTGGGCTCTTGGGAAAAAGCAG 1140
 Qy 1235 GAGTTCTACTCTTTTAAAGGAGCTCTTGGAGTGAACATTCCTCAAGATGCTAATGTGG 1284
 Db 1141 GAGTTCTACTCTTTTAAAGGAGCTCTTGGAGTGAACATTCCTCAAGATGCTAATGTGG 1200
 Qy 1285 TCTTCAAAAGAGCAGAAAGGAGGAGTTCCAGTGAATTTTGGTATTTACATGATGTGAT 1344
 Db 1201 TCTTCAAAAGAGCAGAAAGGAGGAGTTCCAGTGAATTTTGGATTTACATGATGTGAT 1260
 Qy 1345 TAATTAATATCTCATGTGGAATAATATGACATTGAGAAATACAGAAGATGAAGACA 1404
 Db 1261 TAATTAATATCTCATGTGGAATAATATGACATTGAGAAATACAGAAGATGAAGACA 1320
 Qy 1405 ATGACATGGAAGGCTCAGATATAGATGTTCAAGATGATGAATGCGCAAACTGTTTCA 1464
 Db 1321 ATGACATGGAAGGCTCAGATATAGATGTTCAAGATGATGAATGCGCAAACTGTTTCA 1380
 Qy 1465 GAGATGAGGAAGAAATTTACCTTGGAACTGAGTGAAGTGAAGAAACCTTTA 1524
 Db 1381 GAGATGAGGAAGAAATTTACCTTGGAACTGAGTGAAGTGAAGAAACCTTTA 1440
 Qy 1525 ATGCAACAGTATGAGCTTCTGACAGCATAGTACTCTTATGCTGGTTGGCAAGAGAT 1584
 Db 1441 ATGCAACAGTATGAGCTTCTGACAGCATAGTACTCTTATGCTGGTTGGCAAGAGAT 1500
 Qy 1585 CCATGGCATTTTGGCAATCTTATGATGTGGCAGTTAACTGAAA----- 1631
 Db 1501 CCATGGCATTTTGGCAATCTTATGATGTGGCAGTTAACTGAAAAGTTGCTGGGCC 1560
 Qy 1632 ----- 1631
 Db 1561 ATCTTGTCTCTCTTACACAGCAGCATGGCTGGCAGAGAGGAAAGACAGAGCGCCATC 1620
 Qy 1632 -----GGCAGATCTACTATGCTTCTTCTTACTAGAAATA 1662
 Db 1621 TTGGAACGTCATGTCCTTACTTCTGCGGCGACATCTCATGCTTCTTACTAGAAATA 1680
 Qy 1663 ACTGTGCAAGTGGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1722
 Db 1681 ACTGTGCAAGTGGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
 Qy 1723 AGATGTACAGAAGAGCGAGAACCCAGTATCTTATGCTGGAGATGTAGGAACCAAGATC 1782
 Db 1741 AGATGTACAGAAGAGCGAGAACCCAGTATCTTATGCTGGAGATGTAGGAACCAAGATC 1800
 Qy 1783 TCTTAAATTTATCCAGCTCAACAGATTTCTATTCAGTGAATTAACATGATCCAG 1842
 Db 1801 TCTTAAATTTATCCAGCTCAACAGATTTCTATTCAGTGAATTAACATGATCCAG 1860
 Qy 1843 AAGCAGAAGATTTAAGTGGGAAATATATAAGACCTCATCTTGTATTTACTGTGT 1902
 Db 1861 AAGCAGAAGATTTAAGTGGGAAATATATAAGACCTCATCTTGTATTTACTGTGT 1920
 Qy 1903 CAGTATGAGGACTATTTGTCACACATGAAGAAAGCAAAAGATTTTACTGAAGCAG 1962
 Db 1921 CAGTATGAGGACTATTTGTCACACATGAAGAAAGCAAAAGATTTTACTGAAGCAG 1980
 Qy 1963 GAAATACCTAAAGATATGTTATCAGTGAATTTATCTGAAGAAGATGTTTGTCTAC 2022
 Db 1981 GAAATACCTAAAGATATGTTATCAGTGAATTTATCTGAAGAAGATGTTTGTCTAC 2040
 Qy 2023 TGTCAACCAATATGCTGCAAGCTTCCAGCCCTGCTGCTGCCAGACACAGAGAAGCA 2082
 Db 2041 TGTCAACCAATATGCTGCAAGCTTCCAGCCCTGCTGCTGCCAGACACAGAGAAGCA 2100

Qy 2083 AAATAGAGAGCATCCCTAGCTAGACACATGACACAGACATAGTCAAAATATAACAG 2142
 Db 2101 AAATAGAGAGCATCCCTAGCTAGACACATGACACAGACATAGTCAAAATATAACAG 2160
 Qy 2143 ATGCACTACTGGAATAATGTTCCGAAATCACTGTGAAATCTTCCAGTATTTTCAGAC 2202
 Db 2161 ATGCACTACTGGAATAATGTTCCGAAATCACTGTGAAATCTTCCAGTATTTTCAGAC 2220
 Qy 2203 TTGAGAACCATTATGATTTTGTTCAGTATGCGACCTGTAAATCTCAATATAAAG 2262
 Db 2221 TTGAGAACCATTATGATTTTGTTCAGTATGCGACCTGTAAATCTCAATATAAAG 2280
 Qy 2263 CAATATTGACACTGCTGAAGCAGAAATFACTGTGATTCATTTCTCATGCTGGTTAAAC 2322
 Db 2281 CAATATTGACACTGCTGAAGCAGAAATFACTGTGATTCATTTCTCATGCTGGTTAAAC 2340
 Qy 2323 TAAAGAACTCAGTGGGAGAGGAATCTTGGGGCATATTTGATTCCTGCTGCC 2382
 Db 2341 TAAAGAACTCAGTGGGAGAGGAATCTTGGGGCATATTTGATTCCTGCTGCC 2400
 Qy 2383 TTCTCTTCTTGTGTTGGTGAATCTGCAATTCAGTGTGGCCAAATTTTGCATTTCC 2442
 Db 2401 TTCTCTTCTTGTGTTGGTGAATCTGCAATTCAGTGTGGCCAAATTTTGCATTTCC 2460
 Qy 2443 ACCAGGCTATTAATTGAAGAAACCTTGTATGCTGGAAGAAATTAAGAGCAGACTAG 2502
 Db 2461 ACCAGGCTATTAATTGAAGAAACCTTGTATGCTGGAAGAAATTAAGAGCAGACTAG 2520
 Qy 2503 AAAATCATATCAAAATTTTACCTGCTCAAGAAATGAAGAACTCTCTTCCAGCTTATGAT 2562
 Db 2521 AAAATCATATCAAAATTTTACCTGCTCAAGAAATGAAGAACTCTCTTCCAGCTTATGAT 2580
 Qy 2563 TTCTAAGTATGATAGTATGCGGCAACATCTCAACGTGCGACTAGAAAGTTCCCAATGTA 2622
 Db 2581 TTCTAAGTATGATAGTATGCGGCAACATCTCAACGTGCGACTAGAAAGTTCCCAATGTA 2640
 Qy 2623 TGAAGAAACAGATGTCAGAGGAATGATTAAGAAACATGAAATGAATTAATGCGCAGTCA 2682
 Db 2641 TGAAGAAACAGATGTCAGAGGAATGATTAAGAAACATGAAATGAATTAATGCGCAGTCA 2700
 Qy 2683 GAAAGAAACCGATTGAACCTGAGAAATGAAGATTAAGTAAGT 2729
 Db 2701 GAAAGAAACCGATTGAACCTGAGAAATGAAGATTAAGTAAGT 2747

RESULT 4
 AC025659/C
 LOCUS 161100 bp DNA linear HTG 25-JUN-2000
 DEFINITION Homo sapiens clone RP11-263N17, WORKING DRAFT SEQUENCE, 15
 unnumbered pieces.
 ACCESSION AC025659
 VERSION AC025659.3 GI:8705128
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE
 ORGANISM Homo sapiens
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 161100)
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Homo sapiens, clone RP11-263N17
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 161100)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Balwin, J., Barna, N., Bastien, V., Bida, F.,
 Boguslavsky, L., Bouhgalter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Choedel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Plerrie, N., Grant, G., Hago, B., Heaford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Labroque, K., Lamazares, R., Landers, T., Lehoczy, J.,

TITLE
JOURNAL
COMMENT

Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPhee, R.,
Meldrum, J., Meneu, L., Mihova, T., Mihaeva, C., Mienga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Ollivar, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Teste, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (12-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 25, 2000 this sequence version replaced g1:7651991.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center project name: L7940

Center clone name: 263_N_17

Summary Statistics

Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 153679 bases at least Q40
Consensus quality: 157422 bases at least Q30
Consensus quality: 158798 bases at least Q20
Insert size: 163000; agarose-
Insert size: 159700; sum-of-
Quality coverage: 5.1 in Q20 bases; agarose-
Quality coverage: 5.2 in Q20 bases; sum-of-
contigs

NOTE: This is a 'working draft' sequence. It currently
consists of 15 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 3393: contig of 3393 bp in length
3394 3493: gap of 100 bp
3494 8261: contig of 4768 bp in length
8262 8361: gap of 100 bp
8362 13937: contig of 5576 bp in length
13938 14037: gap of 100 bp
14038 23507: contig of 9470 bp in length
23508 23607: gap of 100 bp
23608 30363: contig of 6756 bp in length
30364 30463: gap of 100 bp
30464 39080: contig of 8617 bp in length
39081 39180: gap of 100 bp
39181 47753: contig of 8573 bp in length
47754 47853: gap of 100 bp
47854 59271: contig of 11418 bp in length
59272 59371: gap of 100 bp
59372 71027: contig of 11656 bp in length
71028 71127: gap of 100 bp
71128 84959: contig of 13832 bp in length
84960 85059: gap of 100 bp
85060 95488: contig of 10429 bp in length
95489 95588: gap of 100 bp
95589 108075: contig of 12487 bp in length
108076 108175: gap of 100 bp
108176 123260: contig of 15085 bp in length
123261 123360: gap of 100 bp
123361 140393: contig of 17033 bp in length

* 140394 140493: gap of 100 bp
* 140494 161100: contig of 20607 bp in length.
Location/Qualifiers
1. 161100
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-263N1"
/clone_lib="RP11 Human Male BAC"
1. 3393
/note="assembly-fragment"
3494 8261
/note="assembly-fragment"
8362 13937
/note="assembly-fragment"
14038 23507
/note="assembly-fragment"
23608 30363
/note="assembly-fragment"
vector_end:SP6
clone_side:right
30464 39080
/note="assembly-fragment"
39181 47753
/note="assembly-fragment"
47854 59271
/note="assembly-fragment"
59372 71027
/note="assembly-fragment"
71128 84959
/note="assembly-fragment"
85060 95488
/note="assembly-fragment"
95589 108075
/note="assembly-fragment"
108176 123260
/note="assembly-fragment"
123361 140393
/note="assembly-fragment"
140494 161100
/note="assembly-fragment"

BASE COUNT 47798 a 31689 c 31601 g 48609 t 1403 others
ORIGIN

Query Match 44.1%; Score 1995.8; DB 2; Length 161100;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1997; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2515 CAATTTACCTGCTCAAGATGGAACCTCCTCCAGCTTATGATTTCTAAGTATGA 2514
|||||
Db 114956 CAATTTACCTGCTCAAGATGGAACCTCCTCCAGCTTATGATTTCTAAGTATGA 114897
QY 2575 TAGATGCCGACATCTCAACGTCGACATGGAAGTTCCCAAGTGTATGAAGAACAG 2634
|||||
Db 114896 TAGATGCCGACATCTCAACGTCGACATGGAAGTTCCCAAGTGTATGAAGAACAG 114837
QY 2635 ATGTGCGAGCAATGATTAAGCAACACATGATTAATGCGACATGCAAGAACCGCA 2694
|||||
Db 114836 ATGTGCGAGCAATGATTAAGCAACACATGATTAATGCGACATGCAAGAACCGCA 114777
QY 2695 TTGAACTCTGAGCAATTAACATTTGATGAAGTATGTTAAAGACGAAATAAT 2754
|||||
Db 114776 TTGAACTCTGAGCAATTAACATTTGATGAAGTATGTTAAAGACGAAATAAT 114717
QY 2755 CATTAGACGTGATTAAGATTTAGATGCTCAAAAGTGAACATTAATTTATAGGCTGTG 2814
|||||
Db 114716 CATTAGACGTGATTAAGATTTAGATGCTCAAAAGTGAACATTAATTTATAGGCTGTG 114657
QY 2815 TTTCACAAATTTTTCGATGATAGACTTAATTTATTTCTTAAGCAATATTTAAAT 2874
|||||
Db 114656 TTTCACAAATTTTTCGATGATAGACTTAATTTATTTCTTAAGCAATATTTAAAT 114597
QY 2875 CATTTCAGTTTGACAGCTAGTGCACATCCAAATGAATTAATTAATTAAGTACATATTTTA 2934
|||||

Db 114596 CATTCAGCTTTCGACACTAGTCCCATCCAAATGAAATTAATATTAAGTCACATATTTTA 114537
 Oy 2935 TTTAAATTTTCTAGTACTACTTAACAAGATAAAGTACGAGGGCAAAATTTT 2994
 Db 114536 TTTAAATTTTCTAGTACTACTTAACAAGATAAAGTACGAGGGCAAAATTTT 114477
 Oy 2995 GATTTATCTTTTCACCCAGTAGTATFACCAAAATAGCGAAATATAGAAATTTAAATGAG 3054
 Db 114476 GATATACCTTTTCACCCAGTAGTATFACCAAAATAGCGAAATATAGAAATTTAAATGAG 114417
 Oy 3055 AATTTTACATCTTTTGTGTACCAAGTCTCTAAATGCAGTACATATTTTATAGTACT 3114
 Db 114416 AATTTTACATCTTTTGTGTACCAAGTCTCTAAATGCAGTACATATTTTATAGTACT 114357
 Oy 3115 GCATTTCTACTTCCGAGTAGCCATATTTCAAGTGTTCATGTGCACATGGGCTGTGAC 3174
 Db 114356 GCATTTCTACTTCCGAGTAGCCATATTTCAAGTGTTCATGTGCACATGGGCTGTGAC 114297
 Oy 3175 TACTGTATTTGACAGTTCAGTACTAGACAAAACTAGCATTAATTAATTAGTTCTAGCCA 3234
 Db 114296 TACTGTATTTGACAGTTCAGTACTAGACAAAACTAGCATTAATTAATTAGTTCTAGCCA 114237
 Oy 3235 TGAATTTCTATTTGGATTTAAATTAACCTGAATCAGAGTTAACTCCACAGTGCATTCATG 3294
 Db 114236 TGAATTTCTATTTGGATTTAAATTAACCTGAATCAGAGTTAACTCCACAGTGCATTCATG 114177
 Oy 3295 CAGCTCAGATTAATTTGTTTATTTGAGTCTGATATTAATAAGGCTTGTCAACC 3354
 Db 114176 CAGCTCAGATTAATTTGTTTATTTGAGTCTGATATTAATAAGGCTTGTCAACC 114117
 Oy 3355 TCAGGGGATATTTAGCAATTTGTCGGAGACATTTTGTATGTCAATGAGGCGATTAAT 3414
 Db 114116 TCAGGGGATATTTAGCAATTTGTCGGAGACATTTTGTATGTCAATGAGGCGATTAAT 114057
 Oy 3415 GACATTTAGTGTAGAGGCCATGATCCTGCTAAATTAACCTGATTTGACAGCGCCCA 3474
 Db 114056 GACATTTAGTGTAGAGGCCATGATCCTGCTAAATTAACCTGATTTGACAGCGCCCA 113997
 Oy 3475 CACAAAGAAATTAATCTGCGCCGAATGTAGTGTGCCAAGGCTGATTAACCTTGTGTTA 3534
 Db 113996 CACAAAGAAATTAATCTGCGCCGAATGTAGTGTGCCAAGGCTGATTAACCTTGTGTTA 113937
 Oy 3535 AAAGTAACTGTGGCAGACTAGGTTTCAGAAATTTCTGCTGTCTCAAGTATCATGTT 3594
 Db 113936 AAAGTAACTGTGGCAGACTAGGTTTCAGAAATTTCTGCTGTCTCAAGTATCATGTT 113877
 Oy 3595 TGAATAAATTTTGGCTATTAAAGATATGTATTAGATGCTTATCTGATTAATACCTG 3654
 Db 113876 TGAATAAATTTTGGCTATTAAAGATATGTATTAGATGCTTATCTGATTAATACCTG 113817
 Oy 3655 ATACAACCTGATCTTTCTAATATTTTCAGAAAGTATGAGGATTAACCTTAAGAGAGACT 3714
 Db 113816 ATACAACCTGATCTTTCTAATATTTTCAGAAAGTATGAGGATTAACCTTAAGAGAGACT 113757
 Oy 3715 CAGAATGATATTTATTTTAAAGTGTAGTCTTAAACCTCCTCTTATTTCTACAAGTTATA 3774
 Db 113756 CAGAATGATATTTATTTTAAAGTGTAGTCTTAAACCTCCTCTTATTTCTACAAGTTATA 113697
 Oy 3775 TGGCTAATTTTTCAGATTTGACAGGAGTTCAGATTCCTGCTCATGTGAAGAGAG 3834
 Db 113696 TGGCTAATTTTTCAGATTTGACAGGAGTTCAGATTCCTGCTCATGTGAAGAGAG 113637
 Oy 3835 GGTCCCTCATCTGAAGCGTCTGTGAATATTAACCTTGCAGAGCTTCGACAAATCACTTGA 3894
 Db 113636 GGTCCCTCATCTGAAGCGTCTGTGAATATTAACCTTGCAGAGCTTCGACAAATCACTTGA 113577
 Oy 3895 TCTCCCTGAGCCACAGCGGCTCATTTCTGTAGAGGAGAAAGATTTAGCCAAAGAGTTAAT 3954
 Db 113576 TCTCCCTGAGCCACAGCGGCTCATTTCTGTAGAGGAGAAAGATTTAGCCAAAGAGTTAAT 113517
 Oy 3955 TTTTCATTTCCAAATTCATTAAGTGTAGACTGATCTGTGTGAGCAGTTGTTGTCTCAT 4014
 Db 113516 TTTTCATTTCCAAATTCATTAAGTGTAGACTGATCTGTGTGAGCAGTTGTTGTCTCAT 113457

Oy 4015 TTTGCTGTGACATTTTGTGAGACATTTGTTGAATATTTCTAATTTGGTGTCTACTGTA 4074
 Db 113456 TTTGCTGTGACATTTTGTGAGACATTTGTTGAATATTTCTAATTTGGTGTCTACTGTA 113397
 Oy 4075 TTTTCTTTTAAATCTACTGATATCTGTTCTTTAAATTTTCTTCACATATGTTTG 4134
 Db 113396 TTTTCTTTTAAATCTACTGATATCTGTTCTTTAAATTTTCTTCACATATGTTTG 113337
 Oy 4135 CCTGATACACATGATTTTATTAAGTAAATTTTAAGAAATCTACAGCTAAACTCAGTAA 4194
 Db 113336 CCTGATACACATGATTTTATTAAGTAAATTTTAAGAAATCTACAGCTAAACTCAGTAA 113277
 Oy 4195 GTGCATATTTTCTTATTAACATAGACCCCTGCTACTCTCAGACCCCTCCTCAATTT 4254
 Db 113276 GTGCATATTTTCTTATTAACATAGACCCCTGCTACTCTCAGACCCCTCCTCAATTT 113217
 Oy 4255 TTTTCTCTGATGATGATGCTGATTAACATCAATTTTCAATTTGCTTTTATTTCTAATA 4314
 Db 113216 TTTTCTCTGATGATGATGCTGATTAACATCAATTTTCAATTTGCTTTTATTTCTAATA 113157
 Oy 4315 TGGGAACAAATGAGAGTGAATCTTAATATATAGTTGTATATTAACATCATAGCCTAA 4374
 Db 113156 TGGGAACAAATGAGAGTGAATCTTAATATATAGTTGTATATTAACATCATAGCCTAA 113097
 Oy 4375 TTTATGAATAATCTAATTAATGATACAGACATAGAAACATGAATTTGCTTACTATTTG 4434
 Db 113096 TTTATGAATAATCTAATTAATGATACAGACATAGAAACATGAATTTGCTTACTATTTG 113037
 Oy 4435 ACCTTTGTGACAAATTTTGCAGCTCAATTAATGTTGTCTCATTAATTTAAATAAAGTCTG 4494
 Db 113036 ACCTTTGTGACAAATTTTGCAGCTCAATTAATGTTGTGTATTAATTTAAATAAAGTCTG 112977
 Oy 4495 GGTTCAGATTAATTTTCAAAA 4515
 Db 112976 GGTTCAGATTAATTTTCAAAA 112956

RESULT 5
 CDS1RGC
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 176368)
 Heilig, R., Pelic, J.L., Vico, V., Desilva, C., Robert, C., Winkler, P.,
 Brothier, P., Catolico, L., Barde, V., Pelletier, E., Artiguenave, F.,
 Levy, M., Eckenberger, R., Bruls, T., deBerardinis, V., Cruaud, C.,
 Gysapay, G., Saurin, W. and Weissensbach, J.
 Sequencing of the human chromosome 14
 Unpublished
 2 (bases 1 to 176368)
 Genoscope.
 Direct Submission
 Submitted (08-NOV-2001) Genoscope - Centre National de Sequencage :
 BP 191 91006 Evry cedex - FRANCE (E-mail : segref@genoscope.cns.fr
 On Nov 15, 2001 this sequence version replaced gi:14715170.
 - Web : www.genoscope.cns.fr)
 Center: Genoscope / Centre National de Sequencage
 Genome Center
 Center code: GS
 Web site: http://www.genoscope.cns.fr/
 Contact: Segref@genoscope.cns.fr

COMMENT
 The following BAC sequence is oriented from the T7 to the SP6 end.
 ----- Summary Statistics
 Assembly program: Phrap; version 2.0

Quality coverage: 7.53x in Q20 bases; sum-of-ctrls

Overall quality chart :
Range : bases
0 - 9 : 5
10 - 19 : 25
20 - 29 : 67
30 - 39 : 185
40 - 49 : 2535
50 - 59 : 5835
60 - 69 : 7981
70 - 79 : 20287
80 - 89 : 49616
90 - 99 : 89832

Percentage of bases with a quality value >= 40 : 99 %

FEATURES
Source Location/Qualifiers
1. 176368
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone="R-93311"
/clone_lib="R-93311"
BASE COUNT 49878 a 31876 c 34551 g 60063 t
ORIGIN
Query Match 44.1% Score 1995.8; DB 9; Length 176368;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1997; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 2515 CAATTTTACCTGCTCAAGAAATGGAACCTCCTTCAGCTTATGATTTCTAAGTATGA 2574
Db 154334 CAATTTTACCTGCTCAAGAAATGGAACCTCCTTCAGCTTATGATTTCTAAGTATGA 154393
QY 2575 TAGATGCGGCAACATCAACGCTGACCTAGGAAGTTCCCAAGCTGATGAAAGAACAG 2634
Db 154394 TAGATGCGGCAACATCAACGCTGACCTAGGAAGTTCCCAAGCTGATGAAAGAACAG 154453
QY 2635 ATGTGCGAGGAATGATAGGAACACATGATGAATTAATGCGAGTCAGAAAGAACCGA 2694
Db 154454 ATGTGCGAGGAATGATAGGAACACATGATGAATTAATGCGAGTCAGAAAGAACCGA 154513
QY 2695 TTGAACCTGAGAAATGAAGCATTTGGAATTAATGCTTTAAAGAACAGAAAAAT 2754
Db 154514 TTGAACCTGAGAAATGAAGCATTTGGAATTAATGCTTTAAAGAACAGAAAAAT 154573
QY 2755 CATTAGACGTAAAGATAGGATGCTCAAGAGTCAATTTTATAGGCTGTGG 2814
Db 154574 CATTAGACGTAAAGATAGGATGCTCAAGAGTCAATTTTATAGGCTGTGG 154633
QY 2815 TTTCACAAATTTTGGCATGATAGACTTAATTTCTTAAAGAAATATTTAAAT 2874
Db 154634 TTTCACAAATTTTGGCATGATAGACTTAATTTCTTAAAGAAATATTTAAAT 154693
QY 2875 CATTCAAGTTTGAGACTAGTGCATCCATAGAAATTAATTAATGAACATATTTTA 2934
Db 154694 CATTCAAGTTTGAGACTAGTGCATCCATAGAAATTAATTAATGAACATATTTTA 154753
QY 2935 TTTTAAATTTTCTAGTAACTAATTAACAAAGTAAGGAGGAGGCAAAATATTTT 2994
Db 154754 TTTTAAATTTTCTAGTAACTAATTAACAAAGTAAGGAGGAGGCAAAATATTTT 154813
QY 2995 GATATTACTTTTCAACAGTATATACCAAAATAGCGAATATAGAAATTAATATGAG 3054
Db 154814 GATATTACTTTTCAACAGTATATACCAAAATAGCGAATATAGAAATTAATATGAG 154873
QY 3055 ATATTATACCTCTTTTGTACCAAGCTCTTAATAGCAGTACATATTTTATACTTACT 3114
Db 154874 ATATTATACCTCTTTTGTACCAAGCTCTTAATAGCAGTACATATTTTATACTTACT 154933
QY 3115 GCATTTCTTACTTCGAGTAGACCATATTTCAAGTGTTCATGCCACATGTGGCTGTGAC 3174

|||||
Db 154934 GCATTTCTTACTTCGAGTAGACCATATTTCAAGTGTTCATGCCACATGTGGCTGTGAC 154993
QY 3175 TACTGTATGGACAGTTCAGTACTAGACAAACATAGCATATTAATTAAGTACTAGCA 3234
Db 154994 TACTGTATGGACAGTTCAGTACTAGACAAACATAGCATATTAATTAAGTACTAGCA 155053
QY 3235 TGATTTCTATTTGGATTTAAATTAACCTGATACAGATTAACTCCAGTGTATCATG 3294
Db 155054 TGATTTCTATTTGGATTTAAATTAACCTGATACAGATTAACTCCAGTGTATCATG 155113
QY 3295 CAGCTGACAGTATATTTGTTTATTTGAGTGCATGATATTAATTAATGAGCTTTGTCAAC 3354
Db 155114 CAGCTGACAGTATATTTGTTTATTTGAGTGCATGATATTAATTAATGAGCTTTGTCAAC 155173
QY 3355 TCAGGGATATTTAGCAATTTGTCGGAGACATTTTGTATGATGATGATGAGGAGTATTT 3414
Db 155174 TCAGGGATATTTAGCAATTTGTCGGAGACATTTTGTATGATGATGATGAGGAGTATTT 155233
QY 3415 GACATTTAGTAGTAGAGGCGCATGATCTGTAATTAACCTGCATTTGAGACGCGCCCA 3474
Db 155234 GACATTTAGTAGTAGAGGCGCGAGGATCTGTAATTAACCTGCATTTGAGACGCGCCCA 155293
QY 3475 CAACAAAGATTTATCTGCGCCGAATATGATGCTGCGCAAGGCTGAGTAACTTTGTGTTA 3534
Db 155294 CAACAAAGATTTATCTGCGCCGAATATGATGCTGCGCAAGGCTGAGTAACTTTGTGTTA 155353
QY 3535 AAAGTAACCTGTGCGACACTAGGTTTCCAGAAATTTCTGCTGCTGCTGCTGCTGCTGCT 3594
Db 155354 AAAGTAACCTGTGCGACACTAGGTTTCCAGAAATTTCTGCTGCTGCTGCTGCTGCTGCT 155413
QY 3595 TGAATAATTTTGGCTATTAAGATATGATTTAGATGCTCTTATCCGATTAATTAAGTCTG 3654
Db 155414 TGAATAATTTTGGCTATTAAGATATGATTTAGATGCTCTTATCCGATTAATTAAGTCTG 155473
QY 3655 ATACCACTGATCTTTCTATATATTTTCAAGAGTGTGAGTAACCTCGAAGAGACT 3714
Db 155474 ATACCACTGATCTTTCTATATATTTTCAAGAGTGTGAGTAACCTCGAAGAGACT 155533
QY 3715 CAGAAATATTTATATTTTAAAGTACCTTAAACCTCCTCTTATTTCTACAAGTTATA 3774
Db 155534 CAGAAATATTTATATTTTAAAGTACCTTAAACCTCCTCTTATTTCTACAAGTTATA 155593
QY 3775 TGCGTAAATTTGAGATTTGAACAGGAGATTCAGCATTCGCCATCTCCATGAGAAAGAG 3834
Db 155594 TGCGTAAATTTGAGATTTGAACAGGAGATTCAGCATTCGCCATCTCCATGAGAAAGAG 155653
QY 3835 GCTGCCATCTGAGGAGGCTGCTGAATTCACCTTGAAGCTTCAGACAAATCAGTTGA 3894
Db 155654 GCTGCCATCTGAGGAGGCTGCTGAATTCACCTTGAAGCTTCAGACAAATCAGTTGA 155713
QY 3895 TCTCCCTGAGCCACAGCGCTCATTTCTGTGAGGAGGAGAAATTAAGCCAAAGATTAAT 3954
Db 155714 TCTCCCTGAGCCACAGCGCTCATTTCTGTGAGGAGGAGAAATTAAGCCAAAGATTAAT 155773
QY 3955 TTTCATTTCAAAATCAGTACTGATGATGATCTGTTTGTAGCAGTGTGTTGTCTCAT 4014
Db 155774 TTTCATTTCAAAATCAGTACTGATGATGATCTGTTTGTAGCAGTGTGTTGTCTCAT 155833
QY 4015 TTTGCTGCTGCAATTTTGTAGACATTTGTTGAATATCTATTTGCTGCTGCTGCTGCTGCT 4074
Db 155834 TTTGCTGCTGCAATTTTGTAGACATTTGTTGAATATCTATTTGCTGCTGCTGCTGCTGCT 155893
QY 4075 TTTTCTTTTAAATCTACTGATATCTGTTCTTTAAATTTTCTTCAATATGTTGG 4134
Db 155894 TTTTCTTTTAAATCTACTGATATCTGTTCTTTAAATTTTCTTCAATATGTTGG 155953
QY 4135 CCGATATCAACATATTTTAACTGAATTTAAGATATTAACAGTAAACCTCAGTTAA 4194
Db 155954 CCGATATCAACATATTTTAACTGAATTTAAGATATTAACAGTAAACCTCAGTTAA 156013
QY 4195 GTGCATATTTCTTATTAACATAGACCGCTGTCTACTCTCAGACCCCTCTCTCATATTT 4254
|||||

Db 156014 GTGCAATATATTTCTTATACATAGACCGGTGCTACCTCAGACACCCTCCCTCAATTT 156073
 QY 4255 TTTTTCCTAGCATGTCGATGCGCTGATTAACCTATTTTCATTTGCTTTATTTCTAATA 4314
 Db 156074 TTTTTCCTAGCATGTCGATGCGCTGATTAACCTATTTTCATTTGCTTTATTTCTAATA 156133
 QY 4315 TGGGAACAATGAGAGTGAAGCTCTAATAATAGTTGATAGTAATAAACAATCATTTAGCCTAA 4374
 Db 156134 TGGGAACAATGAGAGTGAAGCTCTAATAATAGTTGATAGTAATAAACAATCATTTAGCCTAA 156193
 QY 4375 TTATTTAGAAAATGCTATTAATAGTACGACACATAGAAAATGCTATGTCATTTGT 4434
 Db 156194 TTATTTAGAAAATGCTATTAATAGTACGACACATAGAAAATGCTATGTCATTTGT 156253
 QY 4435 ACCTTGTCAGCAATTTTGCAGCTCATTAATGTTGTCATTAATTTAAATAAGTCTCG 4494
 Db 156254 ACCTTGTCAGCAATTTTGCAGCTCATTAATGTTGTCATTAATTTAAATAAGTCTCG 156313
 QY 4495 GGTTCAGAAATACCTTCAAAA 4515
 Db 156314 GGTTCAGAAATACCTTCAAAA 156334

RESULT 6
 LOCUS BC027108 1496 bp mRNA linear ROD 07-AUG-2002
 DEFINITION Mus musculus, clone IMAGE:5356411, mRNA, partial cds.
 ACCESSION BC027108
 VERSION BC027108.1 GI:20071983
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (04-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
 COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Jeffrey Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amgdbcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
 Series: IRAC Plate: 54 Row: d Column: 15
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES

source
 1. 1496
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="FVB/N"
 /clone="IMAGE:5356411"
 /tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating ductal carcinoma, 5 month old virgin mouse."
 /clone_lib="NCI_CGAP_Mam6"
 /lab_host="DH10B"
 /note="Vector: pCMV-Sport6"

CDS
 <1..680
 /codon_start=3
 /product="Unknown (protein for IMAGE:5356411)"
 /protein_id="AAH27108.1"
 /db_xref="GI:20071984"
 /translation="PTTLVODMAQILANALLAEPEITVENLPTYLRFORPLLLFSG
 GSINPQYNTLIALVRKQDLSFTFCMINKLNTPYGRGLKAYFQRLPLPOLLVNL
 HSGGVYAFPSOSVTEOSLYLMLHLQAGLENPITYLSAQEMKPLPAFDELNMMDA
 PTSOAPTRKVLVECKEAEVQSEAEIQPDQRSTARRPEVEMLRIRKMNANMKDQEP
 FHHDEKEL"
 BASE COUNT 435 a 329 c 322 g 410 t
 ORIGIN

Query Match 9.4%; Score 426.4; DB 10; Length 1496;
 Best Local Similarity 65.4%; Pred. No. 1,2e-75;
 Matches 756; Conservative 0; Mismatches 326; Indels 74; Gaps 6;
 QY 2101 TAGCTAGCAGACATGACAGACATAGTTCATTAATAACAGATGCTACTGGAATGT 2160
 Db 1 TAGACACACACCTGGTACAAAGACATGGCTCAGATACTAGCAATGCACTAGGAAGCAT 60
 QY 2161 TTCCGGAATTCACGTGTGGAATCTTCCAGTTATTTTCAGACTTCAGAAACATTTTGA 2220
 Db 61 TTCCAGAAATCACTGTGGAATCTTCCAGATATTTAAGATTTTCAGAGACCGCTACTTC 120
 QY 2221 TTTTGTGATGATGCGCATGTAATATCCTCAATATTAATAAGCAATATTCAGACGTGTA 2280
 Db 121 TCTTGTGATGATGCGCATGTAATATCCTCAATATTAATAAGCAATATTCAGACGTGTA 180
 QY 2281 AGCAGAAATCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2340
 Db 181 GACAGAGACAGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240
 QY 2341 GGAAGAGATCTTGGCGGATTTTGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2400
 Db 241 GGAGGGGATCTTGAAGGATTTTGGCGGATTTTGGCGGATTTTGGCGGATTTTGGCGG 300
 QY 2401 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2460
 Db 301 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
 QY 2461 AAAACCTTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2520
 Db 361 AAAAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
 QY 2521 TACCTGCTAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2580
 Db 421 TGTCTGCCAGGATGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 480
 QY 2581 CCGCAACATCTCAAGCTGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2640
 Db 481 CTCTCATATCCCAAGCTGCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
 QY 2641 AGGAGATGATTAAG 2700
 Db 541 AGGAGATGATTAAG 600
 QY 2701 CTCTGATTAAG 2760
 Db 601 TGTGAGATTAAG 660
 QY 2761 GACGTGATTAAG 2820
 Db 661 ACCAG 699
 QY 2821 AAATTTTGGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2880
 Db 700 CTCTCTGTCGACAG 759
 QY 2881 AAGTTTGCAGACTAGTGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2940
 Db 760 ATCTCT--AGACAAGTGCATTGAAGTGCATTGAAGTGCATTGAAGTGCATTGAAGTGC 810

| | | | |
|----|------|--|------|
| QY | 2941 | ATTTCCTGCTAGCAGCATTTAAACAAGTAAGTAGCAGGCGCAAAATATTTGATATT | 3000 |
| Db | 811 | ATTCTGTAGTAGTACTCTTAANAACCGTAAMCCGACGAGA----- | 851 |
| QY | 3001 | ACTTTTCACCCAGTAGTAGTATACCACCAAAATAGCAAAATATAGAAATTTATATGAGATATTT | 3060 |
| Db | 852 | -----TTAACCCCGAGACAGCTTAATTAFTACTACGAAAAATGTTGCATC---- | 891 |
| QY | 3061 | TACATCCTTTTTTTTGACCAAGTCTCTTAATGACATAGATATTTATCTTACCTGATTT | 3120 |
| Db | 892 | ---TTACTTTTTAAGTGAAGTCTTCAAAATATGACATCC--ATTTCACACTTAATGCTTTT | 946 |
| QY | 3121 | CTTACTTCCGAGTAGCCATATTTCAAGTGTCTATTGCCATATGTGGCCCTGTACTACTGT | 3180 |
| Db | 947 | CCTCACTAAGGAGATCCGTTTTTTTGGGCTTAATTTGGCCATAGTGGTACATGACAGTTGT | 1006 |
| QY | 3181 | ATTGGACAGTTTCAGTACTAGACAAAACATAGCATATTAATTAAGTTCTAGCCATGATTT | 3240 |
| Db | 1007 | ATAGGGCAGTTTCACATATGACAGCAACACACAGTGTCTTGACCTGGGTGTAGCCATCTGT | 1066 |
| QY | 3241 | CTATTGGATTAATAAT | 3256 |
| Db | 1067 | GCATTGAATCAAAAT | 1082 |

| RESULT | 7 |
|---|--------------------------------------|
| AC021131 | |
| LOCUS | |
| DEFINITION | |
| AC021131 | 204606 bp DNA linear HTG-07-JUL-2000 |
| Homo sapiens chromosome 14 clone RP11-382120, WORKING DRAFT | |
| SEQUENCE, 27 unordered pieces. | |

| | |
|----------|-------------------------------|
| REVISION | AC021131.5 |
| VERSION | GI:8569986 |
| KEYWORDS | HTG; HTGS_PHASE1; HTGS_DRAFT. |
| SOURCE | Homo sapiens. |
| ORGANISM | Homo sapiens |

REFERENCE
1 (bases 1 to 204606)

AUTHORNS
 TITLE
 JOURNAL
 REFERENCE

WALESTON, R.H.
 The sequence of Homo sapiens clone
 unpublished
 2 (bases 1 to 204606)

AUTHORS
 MALEIKSON, R.H.
 TITLE
 Direct Submission
 Submitted (14-JAN-2000) Genome Sequencing Center, Washington

UNIVERSITY SCHOOL OF MEDICINE, 4444 FOREST PARK PARKWAY, ST. LOUIS
MO 63108, USA
On Jun 16, 2000 this sequence version replaced gi:7534281.

----- Genome Center -----
Center: Washington University Genome Sequencing Center

 Web site: <http://genome.wustl.edu/gsc/index.shtml>

 Project Information

 Contact: E-mail: scott@genome.wustl.edu
 Phone: 314-437-5000
 Fax: 314-437-5000
 Address: 660 South Euclid Avenue
 Box 8080
 St. Louis, MO 63110

```

-----
Contact Project Name: "L19030210"
-----
Summary Statistics
-----
Sequencing vector: M13; 1008
Sequencing vector: M13; 1008
-----

```

| | |
|------------|-------------------------------------|
| Chemistry: | Dye-terminator Big Dye: 0% of reads |
| Chemistry: | Dye-primer ET: 100% of reads |
| Chemistry: | Dye-terminator Big Dye: 0% of reads |
| Chemistry: | Dye-primer ET: 100% of reads |

| Consensus quality: | 192225 bases at least Q40 |
|--------------------|---------------------------|
| Consensus quality: | 195796 bases at least Q30 |
| Consensus quality: | 107516 bases at least Q20 |

Insert size: 209000; agarose- λ p
Insert size: 202006; sum-of-contigs
Qualif. reads used: 450,170,000; bases: 1,350,510,000

* NOTE: This is a working draft; accuracy of coverage: 4.58 in Q20 bases; sum-of-configs

* consists of 27 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

```
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

| | | |
|---|--------|--------------------------------------|
| * | 1 | 1359: contig of 1259 bp in length |
| * | 1260 | 1359: gap of unknown length |
| * | 1360 | 2526: contig of 1167 bp in length |
| * | 2527 | 2526: gap of unknown length |
| * | 2627 | 4233: contig of 1607 bp in length |
| * | 4234 | 4233: gap of unknown length |
| * | 4334 | 6059: contig of 1726 bp in length |
| * | 6060 | 6159: gap of unknown length |
| * | 6160 | 8588: contig of 2429 bp in length |
| * | 8589 | 8688: gap of unknown length |
| * | 8689 | 11532: contig of 2944 bp in length |
| * | 11633 | 11732: gap of unknown length |
| * | 11733 | 14000: contig of 2268 bp in length |
| * | 14001 | 14100: gap of unknown length |
| * | 14101 | 16972: contig of 2872 bp in length |
| * | 16973 | 17072: gap of unknown length |
| * | 17073 | 21089: contig of 4017 bp in length |
| * | 21090 | 21189: gap of unknown length |
| * | 21190 | 25902: contig of 4713 bp in length |
| * | 25903 | 26002: gap of unknown length |
| * | 26003 | 29817: contig of 3815 bp in length |
| * | 29818 | 29917: gap of unknown length |
| * | 29918 | 33616: contig of 3659 bp in length |
| * | 33617 | 33716: gap of unknown length |
| * | 33717 | 37223: contig of 3507 bp in length |
| * | 37224 | 37323: gap of unknown length |
| * | 37324 | 42725: contig of 5402 bp in length |
| * | 42726 | 42825: gap of unknown length |
| * | 42826 | 48145: contig of 5320 bp in length |
| * | 48146 | 48245: gap of unknown length |
| * | 48246 | 52585: contig of 4340 bp in length |
| * | 52586 | 52685: gap of unknown length |
| * | 52686 | 58396: contig of 5711 bp in length |
| * | 58397 | 58497: gap of unknown length |
| * | 58497 | 64865: contig of 6369 bp in length |
| * | 64866 | 64965: gap of unknown length |
| * | 64966 | 70722: contig of 5757 bp in length |
| * | 70723 | 70822: gap of unknown length |
| * | 70823 | 77743: contig of 6921 bp in length |
| * | 77744 | 77843: gap of unknown length |
| * | 77844 | 88107: contig of 10264 bp in length |
| * | 88108 | 88207: gap of unknown length |
| * | 88208 | 99343: contig of 11136 bp in length |
| * | 99344 | 99443: gap of unknown length |
| * | 99444 | 112627: contig of 13184 bp in length |
| * | 112628 | 112727: gap of unknown length |
| * | 112728 | 125006: contig of 12279 bp in length |
| * | 125007 | 125106: gap of unknown length |
| * | 125107 | 136350: contig of 11244 bp in length |
| * | 136351 | 136450: gap of unknown length |
| * | 136451 | 163903: contig of 27453 bp in length |
| * | 163904 | 164003: gap of unknown length |
| * | 164004 | 204606: contig of 40603 bp in length |

```
misc_feature      1. 1259 /note="assembly_name:Contig13
misc_feature      1360. 2526 /note="assembly_name:Contig14
misc_feature      2627. 4233 /note="assembly_name:Contig15
misc_feature      4334. 6059 /note="assembly_name:Contig16
misc_feature      6160. 8588 /note="assembly_name:Contig17
```

```

misc_feature      8689..11632
                   /note="assembly_name:Contig18"
misc_feature      11733..14000
                   /note="assembly_name:Contig19"
misc_feature      14101..16972
                   /note="assembly_name:Contig20"
misc_feature      17073..21089
                   /note="assembly_name:Contig21"
misc_feature      21190..25902
                   /note="assembly_name:Contig22"
misc_feature      26003..29817
                   /note="assembly_name:Contig23"
misc_feature      29918..33616
                   /note="assembly_name:Contig24"
misc_feature      33717..37223
                   /note="assembly_name:Contig25"
misc_feature      37324..42725
                   /note="assembly_name:Contig26"
misc_feature      42826..48145
                   /note="assembly_name:Contig27"
misc_feature      48246..52585
                   /note="assembly_name:Contig28"
misc_feature      52686..58396
                   /note="assembly_name:Contig29"
misc_feature      58497..64865
                   /note="assembly_name:Contig30"
misc_feature      64966..70722
                   /note="assembly_name:Contig31"
misc_feature      70823..77743
                   /note="assembly_name:Contig32"
misc_feature      77844..88107
                   /note="assembly_name:Contig33"
misc_feature      88208..99343
                   /note="assembly_name:Contig34"
misc_feature      99444..112627
                   /note="assembly_name:Contig35"
misc_feature      112728..125006
                   /note="assembly_name:Contig36
clone_end:SP6
vector_side:right"
misc_feature      125107..136350
                   /note="assembly_name:Contig37"
misc_feature      136451..163903
                   /note="assembly_name:Contig38"
misc_feature      164004..204606
                   /note="assembly_name:Contig39"
BASE COUNT      64151 a 38935 c 39288 g 59611 t 2621 others
ORIGIN
Query Match      6.8%; Score 307.2; DB 2; Length 204606;
Best Local Similarity 68.3%; Pred. No. 1.5e-51;
Matches 358; Conservative 0; Mismatches 165; Indels 3; Gaps 1;

```

```

QY 4292 TTCATTTGCTTTTATTTCTAATATGGAACAATGAGTAGAAGTCTAATATAGTTGTA 4351
      |||
DB 29999 TTCATTTGCTTTTATTTCTAATATGGAACAATGAGTAGAAGTCTAATATAGTTGTA 30058
QY 4352 GTAATAAAACATCATTTAGGCTTAATTAAGAAAATGCTAATTAAGTACGACACATAGAA 4411
      |||
DB 30059 GTAATAAAACATCATTTAGGCTTAATTAAGAAAATGCTAATTAAGTACGACACATAGAA 30118
QY 4412 ACATGAATGCTTAGTCAATTTGACCTTTGTCAGCAATTTTACAGTCAATATGTTGT 4471
      |||
DB 30119 ACATGAATGCTTAGTCAATTTGACCTTTGTCAGCAATTTTACAGTCAATATGTTGT 30178
QY 4472 CATTAATTTAATTAAGTGTGCTGGGTTTCAGAAATACCTTCAAA 4515
      |||
DB 30179 CATTAATTTAATTAAGTGTGCTGGGTTTCAGAAATACCTTCAAA 30222

```

```

RESULT 8
LOCUS      AX079596 547 bp DNA linear PAT 22-FEB-2001
DEFINITION Sequence 340 from Patent WO0107611.
ACCESSION  AX079596
VERSION     AX079596.1 GI:13159156
KEYWORDS
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 547)
AUTHORS     Baker, K.P., Goddard, A. and Wood, W.I.
TITLE       Human polypeptides and methods for the use thereof
JOURNALS    Patent: WO 0107611-A 340 01-FEB-2001;
Genentech, Inc. (US)
FEATURES
source      location/Qualifiers
            1..547
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT  155 a 117 c 131 g 139 t 5 others
ORIGIN
Query Match      6.4%; Score 290; DB 6; Length 547;
Best Local Similarity 88.8%; Pred. No. 4e-48;
Matches 348; Conservative 0; Mismatches 37; Indels 7; Gaps 3;

```

```

QY 3995 TAGCAGTTGTTGCTCATTTTGTGCTGTCATTTTGTGACATTTGTGAGATATT 4054
      |||
DB 29699 TATTAATTTGATTAACAAATATTTAAGTCAGATGTCCTATTAATTAATGTTCAAAAT 29758
QY 4055 CTATTTGCTGCTACATGATTTTCTTTAATATCTACTGATAT---CTTGTCTTT 4111
      |||
DB 29759 AAATCTGATCTCTTATACATGATGACTAAAGATGTCATGATATATGCAATGGCTN 29818
QY 4112 AAATTTCTCAATATGTTTGTGCTGATACAACTGATTTTATTAAGTAATTTAAGA 4171
DB 29819 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 29878
QY 4172 ATCTAAGCTAAACCTAGTAAGTGCATMTATTTCTTATAACATAGACCGTTGCTAC 4231
DB 29879 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 29938
QY 4232 TCTCAGACCCCTCCTCAATTTTCTCTAGCATGATGCTGATTTAAATCATAT 4291
DB 29939 TCTCAGACCCCTCCTCAATTTTCTCTAGCATGATGCTGATTTAAATCATAT 29998

```

```

QY 111 CCTCTCCGTCGCGCGCGGCAACCTTTGCGCGCGGCTCGGGAACGCCAGGT 170
      |||
DB 114 CCTCTCTCCGTCGCGCGGCAACCTTTGCGCGCGGCTCGGGAACGCCAGGT 172
QY 171 CCGCGCGCGAGTCCCGGCGAGATACATATGATCATCATGAGAACTTTTGAAGTTG 230
      |||
DB 173 CCGCGCGCGAGTCCCGGCGAGATACATATGATCATCATGAGAACTTTTGAAGTTG 228
QY 231 TTCAAGAAAAATTTGAAGTAGCAAAATAGAAATTAAGAAATTAACAGCATACAGAG 290
DB 229 TTCAAGAAAAATTTGAAGTAGCAAAATAGAAATTAAGAAATTAACAGCATACAGAG 288
QY 291 ACAGCATGGAAGTGTGCTTTAGSAAACAGACAGACAGTAAAAAAACAGCAAAATCC 350
DB 289 --CAGCATGAAGTGTGCTTTAGSAAACAGACAGACAGTAAAAAAACAGCAAAATCC 346
QY 351 GCTCAGATACACATCAGCTGATTAATGTTTCCGCTCAATGCTTTAGAGTTGGAGTC 410
DB 347 GCTCAGATACACATCAGCTGATTAATGTTTCCGCTCAATGCTTTAGAGTTGGAGTC 406
QY 411 TCTTTTGTCATATGTCATTTTTCATCATGCAACAGTAACCTTTTACCAAGTGA 470
DB 407 TCTTTTGTCATATGTCATTTTTCATCATGCAACAGTAACCTTTTACCAAGTGA 466
QY 471 COTCAGAAATATTTAGTACATTTGCAACAGG 502
DB 467 COTCAGAAATATTTAGTACATTTGCAACAGG 498

```


KEYWORDS HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 203610)
AUTHORS Birren, B., Nusbaum, C. and Lander, E.
JOURNAL Mus musculus, clone RP23-248E1
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 203610)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouhgalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B., Choquet, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hages, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., MacDonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testfay, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trifillio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zalnoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 203610)
AUTHORS Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choquet, Y., Collamore, A., Cooke, P., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hages, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Poh, K., Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Testfay, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zalnoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Aug 21, 2002 this sequence version replaced gi:17061665. All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
----- Project Information
Contact: sequence_submissions@genome.wi.mit.edu
Center project name: L19034
Center clone name: 248_E1
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 194595 bases at least Q40
Consensus quality: 198128 bases at least Q30
Consensus quality: 199686 bases at least Q20
Insert size: 20400; agarose-fp
Insert size: 200810; sum-of-contigs
Quality coverage: 6.3 in Q20 bases; agarose-fp
Quality coverage: 6.4 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 29 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1366: contig of 1366 bp in length
1367 1466: gap of 100 bp
1467 2392: contig of 926 bp in length
2393 2492: gap of 100 bp
2493 3944: contig of 1452 bp in length
3945 4044: gap of 100 bp
4045 5485: contig of 1441 bp in length
5486 5585: gap of 100 bp
5586 6815: contig of 1230 bp in length
6816 6915: gap of 100 bp
6916 8703: contig of 1788 bp in length
8704 8803: gap of 100 bp
8804 10621: contig of 1818 bp in length
10622 10721: gap of 100 bp
10722 13000: contig of 2279 bp in length
13001 13100: gap of 100 bp
13101 15018: contig of 1918 bp in length
15019 15118: gap of 100 bp
15119 17273: contig of 2155 bp in length
17274 17373: gap of 100 bp
17374 21030: contig of 3657 bp in length
21031 21130: gap of 100 bp
21131 24393: contig of 3263 bp in length
24394 24493: gap of 100 bp
24494 27284: contig of 2791 bp in length
27285 27384: gap of 100 bp
27385 53071: contig of 25667 bp in length
53072 53171: gap of 100 bp
53172 56584: contig of 3413 bp in length
56585 56684: gap of 100 bp
56685 63948: contig of 7264 bp in length
63949 64048: gap of 100 bp
64049 71335: contig of 7267 bp in length
71336 71435: gap of 100 bp
71436 78031: contig of 6596 bp in length
78032 78131: gap of 100 bp
78132 86041: contig of 7910 bp in length
86042 86141: gap of 100 bp
86142 97779: contig of 11638 bp in length
97780 97879: gap of 100 bp
97880 106798: contig of 8919 bp in length
106799 106898: gap of 100 bp
106899 120969: contig of 14071 bp in length
120970 121069: gap of 100 bp
121070 129692: contig of 8623 bp in length
129693 129792: gap of 100 bp
129793 141332: contig of 11540 bp in length
141333 141432: gap of 100 bp
141433 151350: contig of 9918 bp in length
151351 151450: gap of 100 bp
151451 163581: contig of 12131 bp in length
163582 163681: gap of 100 bp
163682 178518: contig of 14837 bp in length
178519 178618: gap of 100 bp
178619 201081: contig of 22463 bp in length
201082 201181: gap of 100 bp
201182 203610: contig of 2429 bp in length.


```
repeat_region 14206..14251
/note="23 copies 2 mer tc 74% conserved"
repeat_region 15594..15633
/note="20 copies 2 mer gt 90% conserved"
repeat_region 15814..16129
/note="LIMB3 repeat: matches 5851..6181 of consensus"
repeat_region 16868..17005
/note="19 copies 2 mer ca 100% conserved"
misc_feature complement(17752..18167)
/note="match: GSS A0065232"
repeat_region 18176..18830
/note="LIMD3 repeat: matches 7082..7740 of consensus"
repeat_region 19980..20318
/note="MLR1A1 repeat: matches 1..337 of consensus"
repeat_region 20320..20418
/note="LRR12 repeat: matches 572..670 of consensus"
repeat_region 20421..20512
/note="LRR30 repeat: matches 551..643 of consensus"
repeat_region 20513..20736
/note="LRR12 repeat: matches 14..1440 of consensus"
misc_feature complement(20664..20987)
/note="match: GSS B33497"
repeat_region 20807..20985
/note="MIR repeat: matches 65..245 of consensus"
repeat_region 20988..21011
/note="12 copies 2 mer tt 96% conserved"
repeat_region 21201..21393
/note="MIR repeat: matches 57..262 of consensus"
repeat_region 24587..24675
/note="LRR23 repeat: matches 304..395 of consensus"
repeat_region 24902..25190
/note="AluY repeat: matches 1..292 of consensus"
repeat_region 25655..25754
/note="50 copies 2 mer ta 85% conserved"
repeat_region 25991..26294
/note="AluSg repeat: matches 1..312 of consensus"
repeat_region 26461..26661
/note="MIR repeat: matches 19..228 of consensus"
misc_feature 28740..29322
/note="match: GSS A0348225"
repeat_region 29629..29699
/note="MIR repeat: matches 59..131 of consensus"
misc_feature complement(31515..31877)
/note="match: GSS A0100296"
repeat_region 32772..32846
/note="L2 repeat: matches 2675..2750 of consensus"
repeat_region 32825..32951
/note="L2 repeat: matches 1729..1860 of consensus"
repeat_region 33291..33572
/note="Charles repeat: matches 7..291 of consensus"
repeat_region 33577..33745
/note="Charles repeat: matches 581..743 of consensus"
repeat_region 33779..35181
/note="Charles repeat: matches 901..2297 of consensus"
repeat_region 35187..35513
/note="LIMAI10 repeat: matches 5994..6316 of consensus"
misc_feature 35272..35512
/note="match: STS 224589; genomic marker D6S470"
repeat_region 35348..35379
/note="16 copies of ca 100% conserved; differs from
224589"
repeat_region 35522..35782
/note="Charles repeat: matches 2289..2574 of consensus"
repeat_region 35921..35944
/note="12 copies 2 mer aa 96% conserved"
repeat_region 37469..38001
/note="L2 repeat: matches 2089..2627 of consensus"
repeat_region 38068..38375
/note="MER63B repeat: matches 8..335 of consensus"
repeat_region 38376..38542
/note="AluY repeat: matches 133..299 of consensus"
repeat_region 38742..39153
/note="L2 repeat: matches 2310..2705 of consensus"

repeat_region 39184..39447
/note="AluSx repeat: matches 38..302 of consensus"
repeat_region 39451..39832
/note="LIP repeat: matches 558..961 of consensus"
repeat_region 39935..40111
/note="AluSg repeat: matches 132..308 of consensus"
repeat_region 40114..40252
/note="L2 repeat: matches 1555..1693 of consensus"
repeat_region 40486..40575
/note="L2 repeat: matches 2598..2727 of consensus"
repeat_region 41332..41498
/note="MIR repeat: matches 55..228 of consensus"
repeat_region 41558..41862
/note="AluY repeat: matches 1..305 of consensus"
repeat_region 43659..43959
/note="AluSg repeat: matches 1..305 of consensus"
repeat_region 44272..44367
/note="MIR repeat: matches 47..140 of consensus"
repeat_region 45065..45299
/note="MIR repeat: matches 1..258 of consensus"
repeat_region 45488..45925
/note="LIPAT7 repeat: matches 5706..6143 of consensus"
repeat_region 45926..45977
/note="26 copies 2 mer aa 71% conserved"
misc_feature 46301..46562
/note="match: GSS A0167854"
repeat_region 46563..46893
/note="MER2 repeat: matches 1..345 of consensus"
repeat_region 46903..46936
/note="17 copies 2 mer tg 91% conserved"
repeat_region 47565..47793
/note="L2 repeat: matches 1901..2131 of consensus"
repeat_region 48465..48620
/note="MLR1I repeat: matches 245..410 of consensus"
repeat_region 49388..49548
/note="MIR repeat: matches 33..198 of consensus"
repeat_region 49549..49829
/note="TIGER1 repeat: matches 2..273 of consensus"
repeat_region 49830..50123
/note="AluSx repeat: matches 16..309 of consensus"
repeat_region 50124..52184
/note="TIGER1 repeat: matches 273..2417 of consensus"
repeat_region 52185..52247
/note="MIR repeat: matches 198..259 of consensus"
repeat_region 52873..53187
/note="MER2 repeat: matches 15..342 of consensus"
misc_feature complement(52938..53361)
/note="match: GSS A0185926"
repeat_region 54150..54465
/note="LIMAT7 repeat: matches 5963..6288 of consensus"

Query Match 2.7%; Score 124; DB 9; Length 94749;
Best Local Similarity 55.9%; Pred. No. 1.4e-14;
Matches 378; Conservative 0; Mismatches 260; Indels 38; Gaps 6;
```

* and the order in which they appear is completely

TCATGATATTAATAATCAGCGTTTGTCAACCTC 3356

D0 8468 NNNNNNNNNNNNATTTGGTTTATTGACTTC 8436

```
RESULT 15
AC012619/c AC012619 89196 bp DNA linear PRI 31-AUG-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2022K1, complete sequence.
DEFINITION AC012619
ACCESSION AC012619
VERSION AC012619.6 GI:15383776
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 89196)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Unpublished
2 (bases 1 to 89196)
DOE Joint Genome Institute.
Direct Submission
3 (bases 1 to 89196)
DOE Joint Genome Institute and Stanford Human Genome Center.
Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 89196)
DOE Joint Genome Institute and Stanford Human Genome Center.
Submitted (31-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Aug 31, 2001 this sequence version replaced gi:7711556.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.7% of Sequence;
Estimated Total Number of Errors is 0.2.
Location/Qualifiers
1. 89196
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2022K1"
BASE COUNT 30551 a 17310 c 16123 g 25212 t
ORIGIN
Query Match 2.5%; Score 111; DB 9; Length 89196;
Best Local Similarity 64.2%; Pred.No.5.7e-12;
Matches 219; Conservative 0; Mismatches 110; Indels 12; Gaps 3;
QY 2915 AATATAAGTCACATATTTTAAATTTCTAGTACTACTTAACAAGTAAAGT 2974
|||||
Db 20986 AATGTGAGCCACATACATATTTAAATTTCTAGTAGCACATTAACCTAAAGG 20927
|||||
QY 2975 GAGCAGGCGAAATTTTGAATTTACTTTGACCCAGTAGTAGTACCAGTAAATAGCGAA 3034
|||||
Db 20926 AAGCAAGTGAAATTTTAAATTAATTAACCAATATATCGAATATATTGTTCA 20867
|||||
QY 3035 ATATAGA-----AATTATTAAGATATTTTACATCCTTTTGTACCAAGTC 3084
|||||
Db 20866 ACATGTATCAACATTAATAATTTAGTAGATATTTTACAT-TTTTTTCATGCTAAATCT 20808
|||||
QY 3085 TCTAATGCGATGATATTTTACTTACTGATTTCTTACTTCGAGTAGCCATATTTTC 3144
|||||
Db 20807 TTGAATTCGGGTGTATCTTCAATTTACGGCTTATCTCATTTGGACTTGACACATTCG 20748
|||||
QY 3145 AAGTGTTC-ATTGCCACAGTGGCGCTGTGACTAGCTGATTTGACAGTTCAGTAGACA 3203
|||||
Db 20747 AAGTGCTCAATAGCTATGCTAGCTATTGACTACTATTTTGGACAGTGCATTTCTAGACC 20688
|||||
QY 3204 AAACCTAGCATATTAAGTTAGTTCTAGCCATATTTCTAT 3244
|||||
Db 20687 TCAAGACACAGAGTATATTTTATCATGTATGATGAATAT 20647
|||||
```

[illegible]

```

RESULT 2
US-07-906-871-15/c
: Sequence 15 Application US/07906871
: Patent No. 5340739
: GENERAL INFORMATION:
: APPLICANT: Stevens, Richard L.
: APPLICANT: Avraham, Shalom
: TITLE OF INVENTION: HEMATOPOIETIC CELL SPECIFIC
: TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY ELEMENTS OF SERGLIXIN AND USES
: TITLE OF INVENTION: THEREOF
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox
: STREET: 1225 Connecticut Avenue, N.W., Suite 300
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/906,871
: FILING DATE: 19920103
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/816,289
: FILING DATE: 03 JAN 1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/635,544
: FILING DATE: 18-JAN-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US89/03051
: FILING DATE: 13-JUL-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/224,035
: FILING DATE: 13-JUL-1988
: ATTORNEY/AGENT INFORMATION:
: NAME: Cimbalia, Michele A
: REGISTRATION NUMBER: 33,851
: REFERENCE/DOCKET NUMBER: 0627,2830004
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)833-7533
: TELEFAX: (202)833-8716
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 17327 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: both
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
: FEATURE:
: NAME/KEY: exon
: LOCATION: 621..753
: FEATURE:
: NAME/KEY: intron
: LOCATION: 754..9596
: FEATURE:
: NAME/KEY: exon
: LOCATION: 9597..9744
: FEATURE:
: NAME/KEY: intron
: LOCATION: 9745..16396
: FEATURE:
: NAME/KEY: exon
: LOCATION: 16397..17327
: US-07-906-871-15
Query Match 2.0%; Score 91.4; DB 1; Length 17327;
Best Local Similarity 58.9%; Pired. No. 5.3e-13;

```

| Matches | 195: Conservative | 0: Mismatches | 131: Indels | 5: Gaps | 2: |
|---|-------------------|---|-------------|---------|----|
| QY | 2878 | TTCAAGTTTGCAGACTAGTGCATTCGCAATAGAAATTAATTAAGTCACATATTTTATTT | 2937 | | |
| Db | 5964 | TTGTCAGTCTGTACAGCCAACTTTGTCCAAATGAAACCAAAAGCAACACATGTGGAGATT | 5905 | | |
| QY | 2938 | AAAAATTTTCTAGTAACACTAATTAACAAAGTAAAGTGAGCGGCGCAAAATTAATTTGAT | 2997 | | |
| Db | 5904 | TAAATGTTCCAGCGGCACATATTTTAAAGTAAAGAAACAGGCGGAAATTAATTTTATTT | 5845 | | |
| QY | 2998 | A----TTACTTTTCACCCAGTAGTATATCCCAAAATAGCGAAATATAGAAATTAATTA | 3053 | | |
| Db | 5844 | TAACTTAACATATTCOAATATATTGTCTTTAATCATATCATCAATGTAAAGTAATTA | 5785 | | |
| QY | 3054 | GATATTTTATACATCTTTTGTGACCAAGCTCTTCTAAATGACAGTACATATTTTATCTAC | 3113 | | |
| Db | 5784 | GATATTTTTCATCTTTCTTCTCATCTAAGGTTCCAAATCTGTATATTTTGTGATTCAC | 5725 | | |
| QY | 3114 | TGCATTTCTTACTTCCAGTAGTACCAATTTTCAAGTGTTCA--TTGGCACATGTGGCGTGG | 3172 | | |
| Db | 5724 | AGCCTGTCCCAATTTTGCCCTTGCCCACTTTTCAAGTGTTCAAGTGTGACGACAGTGA | 5665 | | |
| QY | 3173 | ACTACTGATTTGACAGTTTCACTACTAGTA | 3203 | | |
| Db | 5664 | GCTACCGTACTGTGACAGACAGACAGCGCTAGCA | 5634 | | |
| RESULT 3 | | | | | |
| US-09-784-316-3 | | | | | |
| ; Sequence 3, Application US/09784316 | | | | | |
| ; Patent No. 6461843 | | | | | |
| ; GENERAL INFORMATION: | | | | | |
| ; APPLICANT: WEI, Ming-Hui et al. | | | | | |
| ; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC | | | | | |
| ; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES | | | | | |
| ; TITLE OF INVENTION: THEREOF | | | | | |
| ; FILE REFERENCE: CLO01139 | | | | | |
| ; CURRENT APPLICATION NUMBER: US/09/784,316 | | | | | |
| ; CURRENT FILING DATE: 2001-02-16 | | | | | |
| ; NUMBER OF SEQ ID NOS: 5 | | | | | |
| ; SOFTWARE: FastSeq for Windows Version 4.0 | | | | | |
| ; SEQ ID NO 3 | | | | | |
| ; LENGTH: 65042 | | | | | |
| ; TYPE: DNA | | | | | |
| ; ORGANISM: Human | | | | | |
| ; FEATURE: | | | | | |
| ; NAME/KEY: misc_feature | | | | | |
| ; LOCATION: (1)...(65042) | | | | | |
| ; OTHER INFORMATION: n = A,T,C or G | | | | | |
| US-09-784-316-3 | | | | | |
| Query Match | | | | | |
| Best Local Similarity 58.8%; Score 74.8; DB 4; Length 65042; | | | | | |
| Matches 188; Conservative 0; Mismatches 122; Indels 10; Gaps 3; | | | | | |
| QY | 2901 | TCCATATAGATATATATATTAAGTACATATTTTAAATTTCTAGTACTACATTA | 2960 | | |
| Db | 22330 | TCCATATAGAAATATATATGTGCTAGCTATACACATATTTTAAATTTCTAGTACCACATTA | 22289 | | |
| QY | 2961 | AACAAGTAAAGTAGAGACGAGCGCAAAATATTTTGATATTAATTTTCAACCCAGTAGATA | 3020 | | |
| Db | 22290 | A--AAAGTAAACAGGTAATTAATTTTACTAAATATTTTATGTAAACCAAAATATTC | 22347 | | |
| QY | 3021 | CCCAAAATAGCGAAATATAGAAATTTAATGAGATATT-----TTACATCCTTTT | 3073 | | |
| Db | 22348 | CAAGACATTATTCATTTCAACATGTAATCAATACAAATAATATGAGATTTTGGCATT | 22407 | | |
| QY | 3074 | GTAACAAGTCTTCTAAATGACAGTACATATTTTATCTTACTGACATTTCTTACTCCGAGT | 3133 | | |
| Db | 22408 | GTTTCGATCCCTCAAAATGCGCAAAATGCAATTTACTTACATCAATCATGATCCAAACT | 22467 | | |
| QY | 3134 | AGGCATATTTCAAGTGT--CATGGCACATGTGGCTGTGACTACTGATTTGAGACAGTTC | 3192 | | |

[illegible]

| | | |
|-----|--------------------|---------|
| 1 | LOCATION: 47291: | unknown |
| 2 | OTHER INFORMATION: | unknown |
| 3 | NAME/KEY: unsure | |
| 4 | LOCATION: 52786 | |
| 5 | OTHER INFORMATION: | unknown |
| 6 | NAME/KEY: unsure | |
| 7 | LOCATION: 52787 | |
| 8 | OTHER INFORMATION: | unknown |
| 9 | NAME/KEY: unsure | |
| 10 | LOCATION: 53384 | |
| 11 | OTHER INFORMATION: | unknown |
| 12 | NAME/KEY: unsure | |
| 13 | LOCATION: 54664 | |
| 14 | OTHER INFORMATION: | unknown |
| 15 | NAME/KEY: unsure | |
| 16 | LOCATION: 59215 | |
| 17 | OTHER INFORMATION: | unknown |
| 18 | NAME/KEY: unsure | |
| 19 | LOCATION: 59235 | |
| 20 | OTHER INFORMATION: | unknown |
| 21 | NAME/KEY: unsure | |
| 22 | LOCATION: 59242 | |
| 23 | OTHER INFORMATION: | unknown |
| 24 | NAME/KEY: unsure | |
| 25 | LOCATION: 63230 | |
| 26 | OTHER INFORMATION: | unknown |
| 27 | NAME/KEY: unsure | |
| 28 | LOCATION: 66614 | |
| 29 | OTHER INFORMATION: | unknown |
| 30 | NAME/KEY: unsure | |
| 31 | LOCATION: 68660 | |
| 32 | OTHER INFORMATION: | unknown |
| 33 | NAME/KEY: unsure | |
| 34 | LOCATION: 68657 | |
| 35 | OTHER INFORMATION: | unknown |
| 36 | NAME/KEY: unsure | |
| 37 | LOCATION: 68718 | |
| 38 | OTHER INFORMATION: | unknown |
| 39 | NAME/KEY: unsure | |
| 40 | LOCATION: 68733 | |
| 41 | OTHER INFORMATION: | unknown |
| 42 | NAME/KEY: unsure | |
| 43 | LOCATION: 68739 | |
| 44 | OTHER INFORMATION: | unknown |
| 45 | NAME/KEY: unsure | |
| 46 | LOCATION: 69785 | |
| 47 | OTHER INFORMATION: | unknown |
| 48 | NAME/KEY: unsure | |
| 49 | LOCATION: 79134 | |
| 50 | OTHER INFORMATION: | unknown |
| 51 | NAME/KEY: unsure | |
| 52 | LOCATION: 79198 | |
| 53 | OTHER INFORMATION: | unknown |
| 54 | NAME/KEY: unsure | |
| 55 | LOCATION: 86336 | |
| 56 | OTHER INFORMATION: | unknown |
| 57 | NAME/KEY: unsure | |
| 58 | LOCATION: 97911 | |
| 59 | OTHER INFORMATION: | unknown |
| 60 | NAME/KEY: unsure | |
| 61 | LOCATION: 97911 | |
| 62 | OTHER INFORMATION: | unknown |
| 63 | NAME/KEY: unsure | |
| 64 | LOCATION: 97911 | |
| 65 | OTHER INFORMATION: | unknown |
| 66 | NAME/KEY: unsure | |
| 67 | LOCATION: 97911 | |
| 68 | OTHER INFORMATION: | unknown |
| 69 | NAME/KEY: unsure | |
| 70 | LOCATION: 97911 | |
| 71 | OTHER INFORMATION: | unknown |
| 72 | NAME/KEY: unsure | |
| 73 | LOCATION: 97911 | |
| 74 | OTHER INFORMATION: | unknown |
| 75 | NAME/KEY: unsure | |
| 76 | LOCATION: 97911 | |
| 77 | OTHER INFORMATION: | unknown |
| 78 | NAME/KEY: unsure | |
| 79 | LOCATION: 97911 | |
| 80 | OTHER INFORMATION: | unknown |
| 81 | NAME/KEY: unsure | |
| 82 | LOCATION: 97911 | |
| 83 | OTHER INFORMATION: | unknown |
| 84 | NAME/KEY: unsure | |
| 85 | LOCATION: 97911 | |
| 86 | OTHER INFORMATION: | unknown |
| 87 | NAME/KEY: unsure | |
| 88 | LOCATION: 97911 | |
| 89 | OTHER INFORMATION: | unknown |
| 90 | NAME/KEY: unsure | |
| 91 | LOCATION: 97911 | |
| 92 | OTHER INFORMATION: | unknown |
| 93 | NAME/KEY: unsure | |
| 94 | LOCATION: 97911 | |
| 95 | OTHER INFORMATION: | unknown |
| 96 | NAME/KEY: unsure | |
| 97 | LOCATION: 97911 | |
| 98 | OTHER INFORMATION: | unknown |
| 99 | NAME/KEY: unsure | |
| 100 | LOCATION: 97911 | |
| 101 | OTHER INFORMATION: | unknown |
| 102 | NAME/KEY: unsure | |
| 103 | LOCATION: 97911 | |
| 104 | OTHER INFORMATION: | unknown |
| 105 | NAME/KEY: unsure | |
| 106 | LOCATION: 97911 | |
| 107 | OTHER INFORMATION: | unknown |
| 108 | NAME/KEY: unsure | |
| 109 | LOCATION: 97911 | |
| 110 | OTHER INFORMATION: | unknown |
| 111 | NAME/KEY: unsure | |
| 112 | LOCATION: 97911 | |
| 113 | OTHER INFORMATION: | unknown |
| 114 | NAME/KEY: unsure | |
| 115 | LOCATION: 97911 | |
| 116 | OTHER INFORMATION: | unknown |
| 117 | NAME/KEY: unsure | |
| 118 | LOCATION: 97911 | |
| 119 | OTHER INFORMATION: | unknown |
| 120 | NAME/KEY: unsure | |
| 121 | LOCATION: 97911 | |
| 122 | OTHER INFORMATION: | unknown |
| 123 | NAME/KEY: unsure | |
| 124 | LOCATION: 97911 | |
| 125 | OTHER INFORMATION: | unknown |
| 126 | NAME/KEY: unsure | |
| 127 | LOCATION: 97911 | |
| 128 | OTHER INFORMATION: | unknown |
| 129 | NAME/KEY: unsure | |
| 130 | LOCATION: 97911 | |
| 131 | OTHER INFORMATION: | unknown |
| 132 | NAME/KEY: unsure | |
| 133 | LOCATION: 97911 | |
| 134 | OTHER INFORMATION: | unknown |
| 135 | NAME/KEY: unsure | |
| 136 | LOCATION: 97911 | |
| 137 | OTHER INFORMATION: | unknown |
| 138 | NAME/KEY: unsure | |
| 139 | LOCATION: 97911 | |
| 140 | OTHER INFORMATION: | unknown |
| 141 | NAME/KEY: unsure | |
| 142 | LOCATION: 97911 | |
| 143 | OTHER INFORMATION: | unknown |
| 144 | NAME/KEY: unsure | |
| 145 | LOCATION: 97911 | |
| 146 | OTHER INFORMATION: | unknown |
| 147 | NAME/KEY: unsure | |
| 148 | LOCATION: 97911 | |
| 149 | OTHER INFORMATION: | unknown |
| 150 | NAME/KEY: unsure | |
| 151 | LOCATION: 97911 | |
| 152 | OTHER INFORMATION: | |

| | | | | |
|-----------------------|--------------|--------------------|----------------|-------------------|
| Query Match | 1.58; | Score 69.2; | DB 4; | Length 87543; |
| Best Local Similarity | 62.7%; | Pred. No. 5.8e-07; | | |
| Matches 148; | Conservative | 0; | Mismatches 73; | Indels 15; Gaps 2 |

[illegible]

[illegible]

| QY | 2964 | Db | 154963 | QY | 2984 | Db | 154909 |
|-----------------|---|-------------------|--|-------------|------|----|--------|
| CACATATATTTT | TTTTTAAAAATTTT | CAGTAAACATCACTTTT | AAAAAAGAAATTAAC | AGTGGACAGGC | 2583 | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| CACATATCTAATTTT | TTAAATTTTTTAAATAGGACACATTTTA | ----- | AAAACTGAAAAAAGA | 154910 | | | |
| QY | 2984 | AAAAATATTTT | GATTTACTTTTACCACAGTAAATCCCAAAATAGCGAAATATAGAAA | 3043 | | | |
| | | | | | | | |
| | | | | | | | |
| AATCTATTTT | AATGATTTTGAATCCAGTGTAACCAAAAAATTTGTTCAACAAGATATTAAT | 154850 | | | | | |

```

OY 3044 TAAATTAACAGATATTATTACATCTTTTGTGACCAAGCTCTCTAATGACATCAATTT 3103
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154849 ATTAATAATTATGAGTTTATTACTTGTATTATTACTAGTCTTTGAAATCTGGTGTATT 154790
OY 3104 TTATCACTACTGCAATTTCTTACTTCGAGTAGACCAATTTTCAAGTG-TTCATTTGCCACAT 3162
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154789 TTACACTTAAGCACACATCATCAGTTTGGATAGACACATTTCCAACTTAAATTAATCTACAT 154730
OY 3163 GTGCCCTGTGACTACTGTAATTGGACAGTTTCAG 3194
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154729 ATGTTAGTGCACTATCTGTGACAGACAG 154698

RESULT 9
US-08-724-394A-22/c
; Sequence 22, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Rudd, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724.394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 246240 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..246240
; OTHER INFORMATION: /note= "HLA-H.CONFIG"
US-08-724-394A-22

Query Match 1.4%; Score 65.2; DB 2; Length 246240;
Best Local Similarity 57.7%; Pred. No. 1.1e-05;
Matches 157; Conservative 0; Mismatches 108; Indels 7; Gaps 2;

OY 2924 CACATATTTTATTTAAATTTCTGTGTAACATTTAAACAAGTAAAGTGACAGGCG 2983
      ||||||| ||| ||| ||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154963 CACATATGCTAAATTTTAAATTTTATATGAGCACATTTT-----AAAGTAAACAAA 154910
OY 2984 AAAATAAATTTTGATTTTACTTTTCACCCAGTGTAGTATACCACAAATATAGCAAAATATAGAAA 3043

```

| | | | | |
|---|----|--------|--|--------|
| | Db | 154909 | AATCTATTTTTAAATGATTTTGAAATCCAGGTGAACCAAAAATTTGTTCCACAGAGTATCTAAT | 154850 |
| | Qy | 3044 | TTATTAAATGAGAATAATTTTACATCCTTTTTTGACCAGTCCTTAATAGCAGTACATAT | 3103 |
| | Db | 154849 | ATTAAATATATTGAGTTTATTACTTCTTATATTACTAGTAGTCTTTGAAATCGTGSTGTAT | 154790 |
| | Qy | 3104 | TTATATCTTACAGCATTTTCTTACTTCCAGTAGGCCATATTTCAAAGT-TTTATTGGCACAT | 3162 |
| | Db | 154789 | TTACACTTAAAGCCACATACAGTTTGGAGTACCACATTTCCAAATGCTTAAATCTACAT | 154730 |
| | Qy | 3163 | GAGGCCTGAGACTGATTTGGACAGATTAC | 3194 |
| | Db | 154729 | ATGTTAGTGGCACTATCTTGGACAGCACAG | 154698 |
| RESULT 10 US-09-484-970B-44 Sequence 44, Application US/09484970B Patent No. 6426186 GENERAL INFORMATION: APPLICANT: Jones, Karen A. APPLICANT: Volkmutz, Wayne APPLICANT: Walker, Michael G. TITLE OF INVENTION: BONE REMODELING GENES FILE REFERENCE: PB-0014 US CURRENT APPLICATION NUMBER: US/09/484,970B CURRENT FILING DATE: 2000-01-18 NUMBER OF SEQ ID NOS: 172 SOFTWARE: PERL Program SEQ ID NO 44 LENGTH: 2821 TYPE: DNA ORGANISM: Homo sapiens FEATURE: NAME/KEY: misc-feature OTHER INFORMATION: Incyte ID No. 6426186 331395.ICB1 US-09-484-970B-44 | | | | |
| Query Match Best Local Similarity 1.4%; Score 63.8; DB 4; Length 2821; Matches 207; Conservative 0; Mismatches 147; Indels 25; Gaps 3; | | | | |
| | Qy | 2853 | TCTTTAAAGAATTAATATTAAATCATTTTCAGTTTGCAGACTAGTCCATCCAAATGAAT | 2912 |
| | Db | 1055 | TTCTTACTAACACAGATTAGGSTATAGGCATCTAGAGATGCACGTCCAGTGAATACT | 1114 |
| | Qy | 2913 | ATATATTAAGTCACATTTTATTATTTAAATTTTCTAGTACTACATTAAACAAGTAAAA | 2972 |
| | Db | 1115 | ATAAAGAGAGGCACACTCTGTAATTTAAATGTTTCTAGTACACACATTTTAAAAAGTAA | 1174 |
| | Qy | 2973 | GTGAGCAGGCGCAAAATATTTTGTGATFATACCTTTTCAC-----CAGTAG | 3016 |
| | Db | 1175 | AAAGAACCCAGGATATGATTTAAATATATTTTATTAAACCAATAGGCTCAAAAACACTAT | 1234 |
| | Qy | 3017 | TATACCCAAATATAGCGAATATAGAAATTTATATAGATATTTTACATCTCTTTTGTGA | 3076 |
| | Db | 1235 | CATTTCACATATAGGATGCATATATAAAATTTATTAATGAGTATTTTACT--TTTTCATA | 1292 |
| | Qy | 3077 | CCAAGTCTTTPAATGACATATTTTATFATCTTACTGCACTTTCTTACTTCCAGTAGAC | 3136 |
| | Db | 1293 | TTAGGCTTTTGAATTTAGTGTGTATFAACA-----CATTCOAATTCAAAACCATCAG | 1345 |
| | Qy | 3137 | CATATTTCAATGTTCATTTGGCACATGTGGCGCTGTGACTACTGATTTTGACAGTTCAGTA | 3196 |
| | Db | 1346 | CATATTTTCAATCTTTTGGAGCCAAATGTGACTAGTGCTCAACATATTAGACAGTGCAGAT | 1405 |
| | Qy | 3197 | CTAGACAAAACATGCAGATA | 3215 |
| | Db | 1406 | CTATATCTCATTTCTTCATA | 1424 |
| RESULT 11 | | | | |

```

1  APPLICANT: Cohen, Annick
2  TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
3  TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
4  FILE REFERENCE: GENSEP.05ICP1
5  CURRENT APPLICATION NUMBER: US/09/641,638
6  CURRENT FILING DATE: 2000-08-16
7  PRIOR APPLICATION NUMBER: US 09/502,330
8  PRIOR FILING DATE: 2000-02-11
9  PRIOR APPLICATION NUMBER: US 60/133,200
10 PRIOR FILING DATE: 1999-05-07
11 PRIOR APPLICATION NUMBER: US 09/275,267
12 PRIOR FILING DATE: 1999-03-23
13 PRIOR APPLICATION NUMBER: US 60/119,917
14 PRIOR FILING DATE: 1999-02-12
15 NUMBER OF SEQ ID NOS: 1304
16 SOFTWARE: Patent.pm
17 SEQ ID NO 24
18 LENGTH: 955
19 TYPE: DNA
20 ORGANISM: Homo Sapiens
21 FEATURE:
22 NAME/KEY: allele
23 LOCATION: 478
24 OTHER INFORMATION: 12-628-311 : polymorphic base T or C
25 NAME/KEY: misc_binding
26 LOCATION: 458..477
27 OTHER INFORMATION: 12-628-311.misl, potential
28 NAME/KEY: misc_binding
29 LOCATION: 479..498
30 OTHER INFORMATION: 12-628-311.mls2, potential complement
31 NAME/KEY: primer_bind
32 LOCATION: 769..787
33 OTHER INFORMATION: upstream amplification primer, complement
34 NAME/KEY: primer_bind
35 LOCATION: 271..291
36 OTHER INFORMATION: downstream amplification primer
37 NAME/KEY: misc_binding
38 LOCATION: 466..490
39 OTHER INFORMATION: 12-628-311 potential probe
40 US-09-641-638-24
41
42 Query Match 1.4%: Score 63.4; DB 4; Length 955;
43 Best Local Similarity 64.9%; Pred. No. 1.5e-06;
44 Matches 111; Conservative 0; Mismatches 56; Indels 4; Gaps 1;
45
46 QY 3043 ATTATTATGAGATATTTCATGCC---TTTTGTACCAAGCTCTTAATGAGATAC 3098
47 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
48 Db 904 ATTATTAAAGCATGTTTGCATTCCTTTCTTTGTTATTAACTCTTCAACACGCCGT 845
49
50 QY 3099 ATATTATTACTTACTGCACTTCTTACTTCCGAGTAGCCATATTTCAGTGTCATGCC 3158
51 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 Db 844 GCAATTATCACTTACAGCAGCGTCTCAGTTGAGAGAGATACACTTCAAGGCGTCAGAGC 785
53
54 QY 3159 ACATGTGGCGCTGTGACTACTGTATTGGACAGTTCAGTACTAGCAAAAAC 3209
55 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
56 Db 784 CACACTGACTAGCAACTGATGCGATTGGGTAGTACACTACAGGCAAAAAT 734
57
58 RESULT 13
59 US-09-813-133A-3/C
60 Sequence 3, Application US/09813133A
61 Patent No. 6455294
62 GENERAL INFORMATION:
63 APPLICANT: GAN, Weiniu et al
64 TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS, AND
65 TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
66 FILE REFERENCE: C1001173
67 CURRENT APPLICATION NUMBER: US/09/813,133A
68 CURRENT FILING DATE: 2001-06-06
69 NUMBER OF SEQ ID NOS: 4
70 SOFTWARE: FastSeq for Windows Version 4.0
71 SEQ ID NO 3

```

```

; LENGTH: 55827
; TYPE: DNA
; ORGANISM: Human
US-09-813-133A-3

```

| | | | | | | | |
|-----------------------|-------|--------------|---------|------------|----|--------|-------|
| Query Match | 1.4% | Score | 62 | DB | 4 | Length | 55827 |
| Best Local Similarity | 70.0% | Pred. No. | 3.1e-05 | | | | |
| Matches | 112 | Conservative | 0 | Mismatches | 45 | Indels | 3 |
| | | | | | | Gaps | 2 |

| | | | |
|----|-------|---|-------|
| Dd | 46475 | AATATAAAATATTCTTGAANAATATTTTAATGTACTTTTTTCATACAGCCTTCAAAAT | 46416 |
| Qy | 3092 | GCAGTACATATTTTATTTACTTACGATTTCTTACTCCAGACGATATTTCAAG-TGT | 3150 |
| Dd | 46415 | CTGGTGCTGATTTTAACTACTCACTGATATGTCTCAATTCAGACAAGCTACTATTTAAGCACT | 46356 |
| Qy | 3151 | TGATTCGCCATGTTGGCGCTGTGACTGACTGATATGGACAGT | 3190 |
| Dd | 46355 | GGATAGCCCATGATGTGTTAGTGTGCTACTACTATGCTGGACAGT | 46316 |

RESULT 14
US-08-370-319C-12/C
; Sequence 12, Application US/08370319C
; Patent No. 5856091

```

1 GENERAL INFORMATION:
2 APPLICANT: Bricbard, Vincent; Van pel, Aline;
3 APPLICANT: Traversari, Catia; W lfel, Thomas; Coulle, Pierre;
4 APPLICANT: Boon-Falleur, Thierry; De phaes, Etienne
5 TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
6 TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR PROCESSED
7 TITLE OF INVENTION: REJECTION ANTIGEN PRESENTED BY HLA-A2

```

US-08-370-319C-12

| | | | | |
|---------------------------|--------|--------------------|------------|--------------|
| Query Match | 1.48: | Score 61.8: | DB 2: | Length 4129: |
| Best Local Similarity | 53.98: | Pred. No. 8,75-06: | | |
| Matches 200, Conservative | 0: | Mismatches 157, | Indels 14, | Gaps 3 |

| | | | |
|----|------|--|------|
| QY | 2926 | CATTTTATTTTAAATTTCTAGTAAAGTAAACAAGTA-----AAAGTACG | 2978 |
| Db | 2849 | CATATGTAAATTTTAAATTTCTAGTACGTACATTTTAAAGAAAGAAACATGTAAATTAAT | 2730 |
| QY | 2979 | AGGGCAAAATAAATTTGCATATTACTTTTGCACCAGTAGTATACCCAAATATGACGAATAT | 3038 |
| Db | 2789 | CTTAATAATTTTGTGTTAAACCTAAATTTTGCAGAAATATTTATTTCAAGGTGAATCAATAT | 2730 |
| QY | 3039 | AGAAATTTATTAATGAGATATTTTACATCCTTTTGTGTTACCAAGTCTTCTAAATGCAGTAC | 3098 |
| Db | 2729 | AAAATTCATTAAGAACAACATTTATCTTCTTTCTCTACATCACTAAATTTTGGAAATATAT | 2670 |
| QY | 3099 | ATATTTATTAAGTACAGCATTTTCTTCTACGCGAGTACGCATATTTCAAGTGTCAATTCG | 3158 |
| Db | 2669 | ATGATTTATTAATCAATG-----CTTAATTTCAAACTCTCTGTGTTTAAGTGTTCAATGACA | 2615 |
| QY | 3159 | ACATGTGGCCTGTGACTACTGTATTTGAGACAGTTTCAGTACTAACACAAATAGCATATTT | 3218 |
| Db | 2614 | CCAGTGTACATGATGGGCTAAACCCATCAGACAGCAGCAGCTATATACATATAAT - ATAGCT | 2557 |
| QY | 3219 | AACCTAGTCTAGCAGCATGATTTCTATTTGAGATTTAAATTTAAATCTTAATCAGAGTTAACT | 3278 |
| Db | 2556 | ACTTTATCCCTCCCTCTGTATATATCTGATTCATCAATTAAGGCAAAATATTAACCTAAAAA | 2497 |
| QY | 3279 | CCACAGTGCAT | 3289 |
| Db | 2496 | CATTAAGAAT | 2486 |

RESULT 15
US-09-224-834-12/c

```

Sequence 12: Application US/09224834
Patent No. 620111
GENERAL INFORMATION:
APPLICANT: Brichard, Vincent; Van pel, Aline;
APPLICANT: Traversari, Catia; W lfel, Thomas; Coule, Pierre;
TITLE OF INVENTION: Boon-Falleur, Thierry; De Plaen, Etienne
TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PERCURSOR PROCESSED TO AT LEAST ONE
TITLE OF INVENTION: REJECTION ANTIGEN PRESENTED BY HLA-A2

```

```

1  REFERENCE/DOCKET NUMBER: LUD 5377.1
2
3  TELECOMMUNICATION INFORMATION:
4  TELEPHONE: (212) 688-9200
5  TELEFAX: (212) 838-3884
6
7  INFORMATION FOR SEQ ID NO: 12:
8
9  SEQUENCE CHARACTERISTICS:
10
11     LENGTH: 4129 base pairs
12     TYPE: nucleic acid
13     STRANDEDNESS: double
14     TOPOLOGY: linear
15
16  FEATURE:
17
18  OTHER INFORMATION: The sequence is preceded by an
19  OTHER INFORMATION: unsequenced portion of from 4.7 to 5.3
20  OTHER INFORMATION: kilobases

```


THIS PAGE BLANK (USPTO)

| | | | |
|----|------|--|------|
| QY | 181 | AGTCCCGGGCAGATACATACATCATCATGATGAGAAACTCTTGTGAAGTTGTTCAAGAAA | 240 |
| Db | 181 | AGGTCCCGGGCAGATACATACATCATCATGATGAGAAACTCTTGTGAAGTTGTTCAAGAAA | 240 |
| QY | 241 | ATTGGAATACCAAAATGAATAATGAATTAACAGACAGATACAGACAGACAGATGGA | 300 |
| Db | 241 | ATTGGAATACCAAAATGAATAATGAATTAACAGACAGATACAGACAGACAGATGGA | 300 |
| QY | 301 | AGTGTGTCCTAGAGAAACAGAACAGACAGATGAGAAAAACAGCAAAATCCGCTCAGATAC | 360 |
| Db | 301 | AGTGTGTCCTAGAGAAACAGAACAGACAGATGAGAAAAACAGCAAAATCCGCTCAGATAC | 360 |
| QY | 361 | AACTGACGCTGATATATGTTTCCGGCTTCAATGCTCTTTAGATTTGGGATCTCTTTGTCA | 420 |
| Db | 361 | AACTGACGCTGATATATGTTTCCGGCTTCAATGCTCTTTAGATTTGGGATCTCTTTGTCA | 420 |
| QY | 421 | TATATGTCATTTTATACATGCCACAGTAACTTTTACAGACATGATGATCTCAGAAAT | 480 |
| Db | 421 | TATATGTCATTTTATACATGCCACAGTAACTTTTACAGACATGATGATCTCAGAAAT | 480 |
| QY | 481 | ATTATAGTACATTTGCAACACAGTCTTTGAAGAACTGATGAGGCTGTAGACCTGTGACG | 540 |
| Db | 481 | ATTATAGTACATTTGCAACACAGTCTTTGAAGAACTGATGAGGCTGTAGACCTGTGACG | 540 |
| QY | 541 | ACTATAGGATTTTCACTGTGCAAGGTTAATGTGTCAAAAGAAATATCAATATCTGTG | 600 |
| Db | 541 | ACTATAGGATTTTCACTGTGCAAGGTTAATGTGTCAAAAGAAATATCAATATCTGTG | 600 |
| QY | 601 | GAAAGAAAGAAAGATTTGATGAAGACATATTTATCAAGGGCAACATATTTGCTCAGAGAT | 660 |
| Db | 601 | GAAAGAAAGAAAGATTTGATGAAGACATATTTATCAAGGGCAACATATTTGCTCAGAGAT | 660 |
| QY | 661 | TCCCTACTGACACCTTTGTTGATGTGAATGCCATTGTGCCCATGTTCTTGTGCTTTC | 720 |
| Db | 661 | TCCCTACTGACACCTTTGTTGATGTGAATGCCATTGTGCCCATGTTCTTGTGCTTTC | 720 |
| QY | 721 | TTTTTATGATGAGAAATTTATACCAACCTGGAAGCCTTCAAGAACATAGAAATGCTC | 780 |
| Db | 721 | TTTTTATGATGAGAAATTTATACCAACCTGGAAGCCTTCAAGAACATAGAAATGCTC | 780 |
| QY | 781 | TGAAAGGAAAGCAAAATATTTATCTCATATGTAAAGCCATTGTGAATACAGAGACA | 840 |
| Db | 781 | TGAAAGGAAAGCAAAATATTTATCTCATATGTAAAGCCATTGTGAATACAGAGACA | 840 |
| QY | 841 | GAGCAGTCAATGGAACCGGTTTTGTGTATGAGGACTACATACCAATTTGTTTAAACACAG | 900 |
| Db | 841 | GAGCAGTCAATGGAACCGGTTTTGTGTATGAGGACTACATACCAATTTGTTTAAACACAG | 900 |
| QY | 901 | AAATGGCCCTTTGGAAAGTATGGCTGTGAGGATGTGGAATATGCAACTCTCTACTTTT | 960 |
| Db | 901 | AAATGGCCCTTTGGAAAGTATGGCTGTGAGGATGTGGAATATGCAACTCTCTACTTTT | 960 |
| QY | 961 | TTTCATTTGTAACATAGTCTTTGCACTTGACCCAGCAATGTAGAAACACATTAATGAACAGC | 1020 |
| Db | 961 | TTTCATTTGTAACATAGTCTTTGCACTTGACCCAGCAATGTAGAAACACATTAATGAACAGC | 1020 |
| QY | 1021 | CATTGACATACATGAACATTTCACTGTTTATTAAGACAATGAAGCACCCTGTGACTG | 1080 |
| Db | 1021 | CATTGACATACATGAACATTTCACTGTTTATTAAGACAATGAAGCACCCTGTGACTG | 1080 |
| QY | 1081 | AGTGTGCAAGATCCTCAACAAGTTTCAAGTGTCCATCTCCAACTGGGCTTACCACTGG | 1140 |
| Db | 1081 | AGTGTGCAAGATCCTCAACAAGTTTCAAGTGTCCATCTCCAACTGGGCTTACCACTGG | 1140 |
| QY | 1141 | TTTTTATTTAGTGAACAGGCTACTTATGAAGCTGATAGAAAGCACTCAGATGGTTG | 1200 |
| Db | 1141 | TTTTTATTTAGTGAACAGGCTACTTATGAAGCTGATAGAAAGCACTCAGATGGTTG | 1200 |
| QY | 1201 | CTTGGCGCTTCTGGGAAAGCAGAGATTTCTACTTTGTTAAAGGACTTTTGAAGTGA | 1260 |
| Db | 1201 | CTTGGCGCTTCTGGGAAAGCAGAGATTTCTACTTTGTTAAAGGACTTTTGAAGTGA | 1260 |

|||||
Db 2341 GGAGAGAACTTGGGGCATTTTGGATCTCTGCTCCCTCTCTCTTTGTTTGG 2400
OY 2401 TGAATCTGATTCAGTGGGCCAAGATTTGGATTTCTTCAACACGCGTATATTAAG 2460
Db 2401 TGAATCTGATTCAGTGGGCCAAGATTTGGATTTCTTCAACACGCGTATATTAAG 2460
OY 2461 AAAACCTGTATTTGGGCGAAGAAATTTGAAGCAGAGCTAGAAAATCATACAAATT 2520
Db 2461 AAAACCTGTATTTGGGCGAAGAAATTTGAAGCAGAGCTAGAAAATCATACAAATT 2520
OY 2521 TACCTGCTCAAGATGGAACCTCTCTTCAGCTTATGATTTTCTAAGTATGATAGATG 2580
Db 2521 TACCTGCTCAAGATGGAACCTCTCTTCAGCTTATGATTTTCTAAGTATGATAGATG 2580
OY 2581 CCGCAACATCTCAAGCTGCGACATAGAAAAGTCCCAAGTGTATGAAGAAACAGATGTGC 2640
Db 2581 CCGCAACATCTCAAGCTGCGACATAGAAAAGTCCCAAGTGTATGAAGAAACAGATGTGC 2640
OY 2641 AGGAAATGATTAAGAACACATGAAGATTAATCGGCACTCAGAAAAGAACCGATTGAA 2700
Db 2641 AGGAAATGATTAAGAACACATGAAGATTAATCGGCACTCAGAAAAGAACCGATTGAA 2700
OY 2701 CTCTGAGAAATTAAGCATTGGAAATGAATTTGTTTAAAGACGAGAAAATCATTTTA 2760
Db 2701 CTCTGAGAAATTAAGCATTGGAAATGAATTTGTTTAAAGACGAGAAAATCATTTTA 2760
OY 2761 GACGCGATTAAGAGATTAGATGCTCAAAAGTAACTAATTTATAGGCTGTGGTTTCCA 2820
Db 2761 GACGCGATTAAGAGATTAGATGCTCAAAAGTAACTAATTTATAGGCTGTGGTTTCCA 2820
OY 2821 AAATTTTGGCATGATAGACTAATTTATTTCTTAAAGATTAATTAATCATTTTC 2880
Db 2821 AAATTTTGGCATGATAGACTAATTTATTTCTTAAAGATTAATTAATCATTTTC 2880
OY 2881 AAGTTTGCAGACTAGTGGCATCAATAGATTAATTAATTAATTAATTAATTAATTA 2940
Db 2881 AAGTTTGCAGACTAGTGGCATCAATAGATTAATTAATTAATTAATTAATTAATTA 2940
OY 2941 ATTTCTGTACTACTAATTAACAAAGTAAAGTGAGCGGCAAAATTAATTTGTTAT 3000
Db 2941 ATTTCTGTACTACTAATTAACAAAGTAAAGTGAGCGGCAAAATTAATTTGTTAT 3000
OY 3001 ACTTTTCAACCCAGTAGTATACCCAAATAGCGAAATATAGAATTAATTAATGAGATTT 3060
Db 3001 ACTTTTCAACCCAGTAGTATACCCAAATAGCGAAATATAGAATTAATTAATGAGATTT 3060
OY 3061 TACATCTTTTGTGTACCAAGCTTCTAATATGACATATTTTATCTACTGCAATT 3120
Db 3061 TACATCTTTTGTGTACCAAGCTTCTAATATGACATATTTTATCTACTGCAATT 3120
OY 3121 CTATCTCGAGTAGCAATTTTCAAGTGTGATGCGCAGATGGCCGTGACATCTCT 3180
Db 3121 CTATCTCGAGTAGCAATTTTCAAGTGTGATGCGCAGATGGCCGTGACATCTCT 3180
OY 3181 ATTGACAGTTCAGTACTAGACAAAACACTAGCATTAATTAATTAATTAATTAATTA 3240
Db 3181 ATTGACAGTTCAGTACTAGACAAAACACTAGCATTAATTAATTAATTAATTAATTA 3240
OY 3241 CTATTTGGATTAATAAATCTAATCACTAATCACTCAAGTGCATTTCACTGACGTG 3300
Db 3241 CTATTTGGATTAATAAATCTAATCACTAATCACTCAAGTGCATTTCACTGACGTG 3300
OY 3301 ACAGTATATTTGTTTATTTGAGTCATGATTTAAATACAGCTTTGCAACCTCAGGG 3360
Db 3301 ACAGTATATTTGTTTATTTGAGTCATGATTTAAATACAGCTTTGCAACCTCAGGG 3360
OY 3361 GATATTTAGCAATTTGCGGAGACATTTTGTATGTCATGACTAGGCACTTTTGCAT 3420
Db 3361 GATATTTAGCAATTTGCGGAGACATTTTGTATGTCATGACTAGGCACTTTTGCAT 3420
OY 3421 TAGTAGTAGAGGCCATGATCTCTGTAATTAATCACTGCACTTGACAGCGCCACACAA 3480
|||||

Db 3421 TAGTAGTAGAGGCCATGATCTCTGTAATTAATCACTGCACTTGACAGCGCCACACAA 3480
OY 3481 AGAATTAATCTGCGCGAATAGTAGTGTGCGCAAGGCTAGTAACCTGTGTTAAAGTA 3540
Db 3481 AGAATTAATCTGCGCGAATAGTAGTGTGCGCAAGGCTAGTAACCTGTGTTAAAGTA 3540
OY 3541 ACCGTGGCAGACTAGTGTCCAGAAATTTCTGTTTCTCTCAGCTATGATGTTTAAAA 3600
Db 3541 ACCGTGGCAGACTAGTGTCCAGAAATTTCTGTTTCTCTCAGCTATGATGTTTAAAA 3600
OY 3601 AATTTTGGCTATTAAGATATGATTAATGATGCTTATATCTGATTAATTAATTAATTA 3660
Db 3601 AATTTTGGCTATTAAGATATGATTAATGATGCTTATATCTGATTAATTAATTAATTA 3660
OY 3661 CTGTGATCTTTCTAATTTTCAAGATGATGCGGTAACCCAGAGAGAGATCTGAAAT 3720
Db 3661 CTGTGATCTTTCTAATTTTCAAGATGATGCGGTAACCCAGAGAGAGATCTGAAAT 3720
OY 3721 GATATTTATATTTTAAAGTATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 3780
Db 3721 GATATTTATATTTTAAAGTATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 3780
OY 3781 AATTTCAATGATTAACAGGATTCAGATTCCTGCAATCTCTCAATGGAAGAGAGCTCC 3840
Db 3781 AATTTCAATGATTAACAGGATTCAGATTCCTGCAATCTCTCAATGGAAGAGAGCTCC 3840
OY 3841 TCATCTGAAGCGCTCTGAAATCTACCTTGAAGCTTCAAGATTAATTAATTAATTAAT 3900
Db 3841 TCATCTGAAGCGCTCTGAAATCTACCTTGAAGCTTCAAGATTAATTAATTAATTAAT 3900
OY 3901 TGAGCCACAGCGCTCTGAGAGGAGGAAATTAAGCAAGATTAATTAATTAATTAATTA 3960
Db 3901 TGAGCCACAGCGCTCTGAGAGGAGGAAATTAAGCAAGATTAATTAATTAATTAATTA 3960
OY 3961 TCCAAATCACTTAAGCTGTTAGACTGATCTGTTTGTAGCAGTTGTTGTCATTTTGGCT 4020
Db 3961 TCCAAATCACTTAAGCTGTTAGACTGATCTGTTTGTAGCAGTTGTTGTCATTTTGGCT 4020
OY 4021 CTGTGATTTTGTAGACATTTGTTGAGATATTTCAATTTGTTGTTGTTGTTGTTGTTG 4080
Db 4021 CTGTGATTTTGTAGACATTTGTTGAGATATTTCAATTTGTTGTTGTTGTTGTTGTTG 4080
OY 4081 TTTTAAATCACTGATATCTGTTTAAATTTTCTCAATATGTTTGGCTGAT 4140
Db 4081 TTTTAAATCACTGATATCTGTTTAAATTTTCTCAATATGTTTGGCTGAT 4140
OY 4141 ACACAGATTTTATTAACGAATTTAAGGAATCTAAGCTAATAAATCACTAGTAGTCAAT 4200
Db 4141 ACACAGATTTTATTAACGAATTTAAGGAATCTAAGCTAATAAATCACTAGTAGTCAAT 4200
OY 4201 MTATTTGCTTATTAACATGACCGGTTGCTACCTGACACCCCTCTCAATTTTCTTC 4260
Db 4201 MTATTTGCTTATTAACATGACCGGTTGCTACCTGACACCCCTCTCAATTTTCTTC 4260
OY 4261 CTGTAGCATGATGATGCTGATTAACCTCAATTTTCAATTTTCAATTTTCAATTTTCAAT 4320
Db 4261 CTGTAGCATGATGATGCTGATTAACCTCAATTTTCAATTTTCAATTTTCAATTTTCAAT 4320
OY 4321 CAATGAGATGAACCTTAATATAGTGTGATTAATAAATCACTATTAATTAATTAATTA 4380
Db 4321 CAATGAGATGAACCTTAATATAGTGTGATTAATAAATCACTATTAATTAATTAATTA 4380
OY 4381 GAAATGCTAATTAATTAACAGACATTAAGCAATTAATTAATTAATTAATTAATTAATTA 4440
Db 4381 GAAATGCTAATTAATTAACAGACATTAAGCAATTAATTAATTAATTAATTAATTAATTA 4440
OY 4441 GTGAGCAATTTTGACAGTCAATTAATGTTTGTCTAATAATTTAATAAAGTGTGCGTTTC 4500
Db 4441 GTGAGCAATTTTGACAGTCAATTAATGTTTGTCTAATAATTTAATAAAGTGTGCGTTTC 4500
OY 4501 AGAATACCTTCAAAAAA 4526
Db 4501 AGAATACCTTCAAAAAA 4526

```

? APPLICANT: Maemura, Koji
? APPLICANT: Hsieh, Chung-Ming
? TITLE OF INVENTION: METHODS OF MODULATING OF ANGIOGENESIS
? FILE REFERENCE: 05433/037001
? CURRENT APPLICATION NUMBER: US/10/121,235
? CURRENT FILING DATE: 2002-04-12
? PRIOR APPLICATION NUMBER: US 09/374,454
? PRIOR FILING DATE: 1999-08-13
? PRIOR APPLICATION NUMBER: US 60/096,515
? PRIOR FILING DATE: 1998-08-14
? NUMBER OF SEQ ID NOS: 22
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 20
? LENGTH: 6792
? TYPE: DNA
? ORGANISM: Homo sapiens
US-10-121-235-20

Query Match          2.2%   Score 97.8:   DB 9:   Length 6792:
Best Local Similarity 65.2%:   Pred No.1.9e-11:
Matches 208:   Conservative 0:   Mismatches 102:   Indels 9:   Gaps 4

? 2889 AGACTAGTGCCCTCCAAATAGATTAATATATAGTCACATATTTATTTAAATTTCTA 2948
? ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
? Db 4797 AGAGCAGGCGCTCCAAATAGAAATTAATCTGACGACACATGATATATT -TATTTCT 4854
? 2949 GTACTACATTAAACAAGTAAGTAAGTGAGCAGGCAAAA---TAATTTGATTAATCT 3003
? Db 4855 CTAGCCACATTTAAAGAAAGTAAGATACAACTAGAACATAATTTAATTTTAATCTG 4914
? 3004 TTTCACCCAGTAGTAGTACCACAAATAGGCAAAATATAGAAATATTAATAGATATTTTC 3063
? ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
? Db 4915 TATATCCAAATATATATTGTAACATGATTAATTAATTAATTAATTAATGATATTTTC 4974
? 3064 ATCCCTTT-TTGATCCAAAGCTCTTAATATGACATATATTTTATCTACTACATTTCT 3122
? ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
? Db 4975 ATTCTTTGGTATATCTAGTCTTCAAAATCTGTAATGTAATCTTACATGATAGCAACAT 5034
? ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
? 3123 TACTTCGAGTAGCCATATTTCAAGTGTTCA-TTGCCACATGTGGCTGTACTACTGTA 3181
? ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
? Db 5035 CACTTTGTACTAGACACATTTGCATGCTCAGTAGCCACACATGTGGCTAGTCTACTGCA 5094
? ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
? 3182 TTGCAGCTTTCAGTACTAG 3200
? ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
? Db 5095 CTGGACAGCAGACGTTCTAG 5113
? ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
US-09-852-067-3/C
? Sequence 3, Application US/09852067
? Patent No. US20020076777A1
? GENERAL INFORMATION:
? APPLICANT: MERKULOV, Gennady et al
? TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
? TITLE OF INVENTION: PROTEIN, NUCLEIC ACID MOLECULES ENCODING HUMAN
? TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS,
? TITLE OF INVENTION: AND USES THEREOF
? FILE REFERENCE: CLO000897-CIP
? CURRENT APPLICATION NUMBER: US/09/852,067
? CURRENT FILING DATE: 2001-05-10
? NUMBER OF SEQ ID NOS: 4
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 3
? LENGTH: 31208
? TYPE: DNA
? ORGANISM: Human
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)...(31208)
? OTHER INFORMATION: n = A,T,C or G
US-09-852-067-3

Query Match          1.9%   Score 86.8:   DB 10:   Length 31208:

```



```

; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11995
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC021880.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.68
; US-09-864-761-11995
```

```

Query Match      1.8%; Score 79.8; DB 10; Length 546;
Best Local Similarity 60.2%; Pred. No. 5.2e-08;
Matches 210; Conservative 0; Mismatches 122; Indels 17; Gaps 4;

QY 2859 AAGAAATATTTAAATCATTTCAGTTTGCAGTACGTCATCCAAATAGATTTAATA 2918
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 202 AAATTTTATCCAAATCAACCAATCTAGATCATGCTGCTCCATAGAAATACATG 261
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2919 TAAGTACATATTTTATTTAAATTTTCTAGTACATCA-TTAAACAAGTAAAGTACG 2977
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 262 CAAGCCACATGTTTAAACTTAATTTCTAGTACCATTTTAAAAAGTAAACAGGT 321
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2978 CAGGCAAAATTAATTTGATATTAATTTTCCACCAGTAGATA-----CCCAA 3026
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 322 GAATCATATGTTAATATATATTTTAAACCAATATATCAATATATATATGACTTCAACAT 381
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3027 ATAGCGAATATAGAAATTTTATGATGATATTTTACATCTTTTGTACCAAGTCTTC 3086
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 382 ATATATCATCAAAAATAGTAGTAGATACGTATATTTTTCATACATAAGTCTTC 441
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3087 TAAATGCGATACATTTTATTTACTTACTGATTTTCTTACTTCCGAGTAGCCATTTTCAA 3146
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 442 AAATCCAGTGTGCTTTTACACTTAGAGCAACATCTCAATGAGGA----CCTCATTTCAA 497
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3147 GTGTTCATTG-CCACAGTGTGCGCTGTGACATGATGATGACAGCTTACG 3194
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

DB 498 GTGCTCAGCAGCATGTGTGCTGTGCGCTGCTGATTTGACACAGCAG 546
RESULT 8
US-09-742-312-3/C
; Sequence 3, Application US/09742312
; Patent No. US20020045166A1
; GENERAL INFORMATION:
; APPLICANT: CHANDRAMOULISARAN, Ishwar et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; FILE REFERENCE: CL000838
; CURRENT APPLICATION NUMBER: US/09/742,312
; CURRENT FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 147309
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(147309)
; OTHER INFORMATION: n = A,T,C or G
; US-09-742-312-3
```

```

Query Match      1.8%; Score 79.6; DB 10; Length 147309;
Best Local Similarity 61.8%; Pred. No. 8.2e-07;
Matches 183; Conservative 0; Mismatches 99; Indels 14; Gaps 3;

QY 2913 ATATATATAGTCACATATTTTATTTAAATTTTCTAGTACATCAATTAACAAGTAAAA 2972
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 132044 AAAAGATGAGCCACATGTAGTATTTCAATTTTCTACTACCCCATCTTAAAGTAAAA 131985
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2973 GTGAGCAGGCGAAATATTTTGTATTTACTTTTACCCAGTAG-----TATAC 3021
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 131984 TTGAACAGGTGAATTAATTTTGTATATATTAATGAAGCAGATCAAAAATATCATTTT 131925
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3022 CCAAAATAGCGAAATATAGAAATTTATATATATATTTTATATTTTATATTTTGTACCAAG 3081
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 131924 CACATGTATAGCATTTTAAACATATATTAAGATAGCTTACCGATTTTTCATATACAG 131865
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3082 TCTTCAATATGACATATTTTATATTTTACTTCTTACTTCTTCCGAGTACATAT 3141
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 131864 TCTTCAAACTCCAGTGTGAATTTTATCTTACAGACATCT--ATTAGACCGCCACAT 131807
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3142 TTCAAGTG-TTCATTTGCGACATGTGCGCTGTGACTGATGATGACAGTTCAGTA 3196
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 131806 TTTAAGTGTGTGATAGGCGTATGTGTGTAGTGTGCTACACACTGAGACAGCAGTA 131751
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

RESULT 9
US-09-741-149-3
; Sequence 3, Application US/09741149
; Patent No. US20020031800A1
; GENERAL INFORMATION:
; APPLICANT: LI, Zhenya et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000780
; CURRENT APPLICATION NUMBER: US/09/741,149
; CURRENT FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 34337
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(34337)
```



```

; TITLE OF INVENTION: DISORDERS SUCH AS SCHIZOPHRENIA
; FILE REFERENCE: 3322/1H702 US1
; CURRENT APPLICATION NUMBER: US/09/935,464
; CURRENT FILING DATE: 2001-08-23
; PRIOR APPLICATION NUMBER: US 09/757,300
; PRIOR FILING DATE: 2001-01-09
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 157875
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-935-464-1

Query Match          1.6%; Score 74; DB 9; Length 157875;
Best Local Similarity 59.4%; Pred. No. 1.4e-05;
Matches 190; Conservative 0; Mismatches 110; Indels 20; Gaps 3;

QY 2889 AGACTAGTCCCATCCATATGATTAATATATAGTCACATATTTATTTAAATTTTCTA 2948
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 88230 AGAGGAGTGCCAGCCAGCAATATAAGAAAGGAGCCATAGGCAATCGTGGTTTCTA 88309
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2949 GTAATCATTAACAAGTAA-----GTAGCAGGCGCAAAATATTTTGG 2995
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 88310 GTAGCCACAGTAAAAAAGTGAAAGAGAGTGATGATCATGATTAATACCATATTTT 88369
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2996 ATATTACTTTTCCACCAG-----TAGTATACCCAAATATAGCAATATATGAAATTTATTA 3050
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 88370 ATAAATTCATCTATCCAGAAATGTATCACTCACTCACTGAAATCAATATATAATGATTTAA 88429
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3051 TGAGATTTTTCATCCCTTTTGTACCAAGTCTCTAATGACAGTACATATTTTACT 3110
    ||| ||||| ||||| ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| |||
DB 88430 TGAGATGATTTACATTTTTCATCAACAGTCTTCAATCTGGTGTATTTCCACT 88489
    ||| ||||| ||||| ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3111 TACTCATTTCTTACTTCCAGTAGCA--TATTTCAAGTGTTCATGCGACATGTGGCC 3168
    ||| ||||| ||||| ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| |||
DB 88490 TACAGACACTCCAGTGTGAGCTTGGCAGTTTTCAGAGTCTGCAGAGTGGCATGTGGCT 88549
    ||| ||||| ||||| ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3169 TGTGACTACTGTATTGGACA 3188
    ||| ||||| ||||| ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| |||
DB 88550 AATGGCTACATATAGTGAAA 88569
    ||| ||||| ||||| ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-09-736-960-86/c
; Sequence 86, Application US/09736960
; Patent No. US20020102267A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter
; APPLICANT: Garman, Jonathan David
; APPLICANT: Candia III, Albert Frederick
; APPLICANT: Airdor Vita Corporation
; TITLE OF INVENTION: CLASP-5 Transmembrane Protein
; FILE REFERENCE: 020054-00051IUS
; CURRENT APPLICATION NUMBER: US/09/736,960
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: US 60/160,860
; PRIOR FILING DATE: 1999-10-21
; PRIOR APPLICATION NUMBER: US 60/162,498
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 60/170,453
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 60/176,195
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 60/182,296
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: US 09/547,276
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,267
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,527
```

```

; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,528
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 09/687,837
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,503
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,508
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,539
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,543
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 134
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 86
; LENGTH: 66686
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: ordered human genomic DNA at CLASP-5 locus
US-09-736-960-86

Query Match          1.6%; Score 72.2; DB 10; Length 66686;
Best Local Similarity 59.7%; Pred. No. 2.4e-05;
Matches 181; Conservative 0; Mismatches 108; Indels 14; Gaps 3;

QY 2913 ATATATTAAGTCACATATTTTATTAATAATTTCTAGTACATTAACAAGTAA 2972
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 24060 ACAATGGGGCCACATATGCTATTTATATATGTTGTAACCAATTAATAAGTAAAGA 24001
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2973 GTGAGCAGGCGCAAAATATTTGATATTTACTTTTACCCAGTAGTATACCAAAATAGC- 3031
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 24000 AAAAGTGAATATTTTATATATTTTGTATACCTAAATAATAATAATCTGCTCC 23941
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3032 -----GAATATTAAGTAAATTTATTAAGATATTTTACATCTTTTGTACCAAGTCTT 3085
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23940 ATGTAGCAATATTAATAATTTACTAGTGAATATTTTATCTTTTATACATTAACCTT 23881
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3086 CTAAATGACATATATTTTATCTT-----ACTGATTTCTTACTTCCGAGTAGCCAT 3139
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23880 CAAATCTGTGCGATTTTTCATCTATCTGTATATATATTCATTTGAAGTAGCCAT 23821
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3140 ATTTCAG-TGTTCAATGGCAGATGTGGCTGTGACTACTGATATGAGAGTTCAGTACT 3198
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 23820 ATTTCAAGATTGATGACATGTGCTAATGCTACTGATATGAGACAGCAAAATCT 23761
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3199 AGA 3201
    |||
DB 23760 AGA 23758

RESULT 14
US-09-822-246-3/c
; Sequence 3, Application US/09822246
; Patent No. US20020142383A1
; GENERAL INFORMATION:
; APPLICANT: MERKLOV, Gennady et al.
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: C1001149
; CURRENT APPLICATION NUMBER: US/09/822,246
; CURRENT FILING DATE: 2001-04-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 197997
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(197997)
```


OTHER INFORMATION: n = A,T,C or G
US-09-822-246-3

| | | | | |
|---------------------------|-------|-----------------|------------|----------------|
| Query Match | 1.6% | Score 71: | DB 10: | Length 197997: |
| Best Local Similarity | 56.4% | Pred. No. | 7.3e-05: | |
| Matches 216; Conservative | 0; | Mismatches 155; | Indels 12; | Gaps 4; |

| | | | |
|----|-------|---|-------|
| QY | 2913 | ATTATATTAGTCACATATTTTATTTTAAAAATTTCTAGTACATTAATTAACAGTAA | 2972 |
| Db | 74207 | AAAAATGTAGCCACGATATGTATTTTAAATTTTCCAGAGAGACATTTAAACAAACAA | 74148 |
| QY | 2973 | GTGACGAGGCGCAAAATATTTTGATATTACTTTTCACCGAGTAGTATACCAA-----A | 3026 |
| Db | 74147 | TCACGTAAGAATTAATTTTAAATATAGTACATTAAACCAATATATGTAAATATATT | 74088 |
| QY | 3027 | ATAGCGAAATATAGAAATATTATTATAGCATATTTTACATCCTTT--TTTGTACCAAGTCT | 3084 |
| Db | 74087 | AACGTGTATCATCAATTAATAATATTATGAGATATTTTGCAATTCCTTTATATAGTCTAGTCT | 74028 |
| QY | 3085 | TCTAATATGCA---GTACATATTTTATTACTACTGATTTCTTCACTCCGAGTAGCATAT | 3141 |
| Db | 74027 | TTGAATCCACTGCTATGTATTTTAAATTTTAAACCAACATCTCAACTGGACTGCTACAT | 73968 |
| QY | 3142 | TTCAAGTG--TTCATTTGCCACATGTGGCCCTGTACTACTGTATTGGACAGTCTAGTACTAG | 3200 |
| Db | 73967 | TTTCAAGTCTGTTGTATGTACATATGAAGCTATAGTGTACCATGTGGACAAACGACGCTTAA | 73908 |
| QY | 3201 | ACAAAACTATGACATATTTAACTTAGTCTTACCGCATGATTTCTATTTGGATTAAAAATTAA | 3260 |
| Db | 73907 | GAAAAACACTGACACACTGTTTACATTTTGGGTTTATATTTATTTTGGATTTAAAGTAAAT | 73848 |
| QY | 3261 | CTCTATATGCAGTTAACTCCACA | 3283 |
| Db | 73847 | CTTGACTGTTGCCCTCATTTTACAA | 73825 |

```

RESULT 15
US-09-764-847-1297
: Sequence 1297, Application US/09764847
: Patent No. US20020132767A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PC009
: CURRENT APPLICATION NUMBER: US/09/764,847
: PRIORITY FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 2003
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1297
: LENGTH: 12047
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-764-847-1297

```

| | | | | |
|-----------------------|--------------|-----------|-----------------|---------------|
| Query Match | 1.5% | Score 70: | DB 10: | Length 12047; |
| Best Local Similarity | 59.2% | Pred. No. | 3.2e-05; | |
| Matches 174; | Conservative | 0; | Mismatches 115; | Indels 5; |
| | | | | Gaps 3; |

[illegible]

Db 1527 GAATTTTGGTGACATATTTTACACTTAAAGGACATCTGCATTTAGACACTCCACCATTTCAA 1586

Qy 3147 GTGTTCA-TTGCCACATGTGGCCTGTGACTACTGATTTGGACAGTTTCAGTACTA 3199

Db 1587 GTGCTCAGTGGCTGCATGTGCTCGTGTGGGTACACATATTTGGACAGCAAGCTCTA 1640

```
Search completed: March 22, 2003, 14:37:45
Job time : 2155 secs
```

THIS PAGE BLANK (USPTO)

Dh 181 TGCCAGGTTAATTGTGTCAAAAGAAATATACAGATCTGTGGAAAAAGAAAGATTT 240
Qy 617 GATGAAGCATATTTATTTCAAGGGCAACATATTTGCTGAGAAATTCCTTACTGACACTT 676
Dh 241 GATGAAGCATATTTATTTCAAGGGCAACATATTTGCTGAGAAATTCCTTACTGACACTT 300
Qy 677 GTTTGATGTGAATGCCATTTGTGCCCATGTTCTTTGCTCTCTTTTAACTGAAGTGA 736
Dh 301 GTTTGATGTGAATGCCATTTGTGCCCATGTTCTTTGCTCTCTTTTAACTGAAGTGA 360
Qy 737 ATATATATACCAACCTGGAAGGCTTCAGACATAGAAAATTCCTGGAAGGAAAAAGCAA 796
Dh 361 ATATATATACCAACCTGGAAGGCTTCAGACATAGAAAATTCCTGGAAGGAAAAAGCAA 420
Qy 797 TATATATATTCATATGTAAAGGCCATTGGAATACAGAGACAGAGCATGTAAGC 856
Dh 421 TATATATATTCATATGTAAAGGCCATTGGAATACAGAGACAGAGCATGTAAGC 480
Qy 857 CGGTTTGTGTATGGGACATACCAATTTGTCTTAACCAAGAAATTCCTTTTGA 916
Dh 481 CGGTTTGTGTATGGGACATACCAATTTGTCTTAACCAAGAAATTCCTTTTGA 540
Qy 917 AAGTATGGGCTCTGAGAGTGGATATGACATCTCTACTTTTTCATTTGAACATAGT 976
Dh 541 AAGTATGGGCTCTGAGAGTGGATATGACATCTCTACTTTTTCATTTGAACATAGT 600
Qy 977 CTGTGACTTGACCCAGCAATGTAGAGAACAATAATGGAACGCCATTGACTACACTGA 1036
Dh 601 CTGTGACTTGACCCAGCAATGTAGAGAACAATAATGGAACGCCATTGACTACACTGA 660
Qy 1037 CATTCACCTGTTTATTAAGAACAAATGAAGCACTCTGTGACTGAAGTGTGAAGATCC 1096
Dh 661 CATTCACCTGTTTATTAAGAACAAATGAAGCACTCTGTGACTGAAGTGTGAAGATCC 720
Qy 1097 TCAACAAGTTTCAACCTCTCCATCTCAACTGGGCTTACCACTGGTTTATTTGTTAGCA 1156
Dh 721 TCAACAAGTTTCAACCTCTCCATCTCAACTGGGCTTACCACTGGTTTATTTGTTAGCA 780
Qy 1157 ACAGGCTACTTANGAGCTGTATAGAGAACTGCAGAAATGGTGTCTTGGCTTCTGGG 1216
Dh 781 ACAGGCTACTTANGAGCTGTATAGAGAACTGCAGAAATGGTGTCTTGGCTTCTGGG 840
Qy 1217 AAAAGCAGAGTCTACTCTGTTAAAGGAACTCTTTGGAAGTGAACATTCCTCAAGATGC 1276
Dh 841 AAAAGCAGAGTCTACTCTGTTAAAGGAACTCTTTGGAAGTGAACATTCCTCAAGATGC 900
Qy 1277 TAAATGGTCTTCAAAAAGAGCAGAGAAGGAGTCCAGTGGAAATTTTGTATTAACATGA 1336
Dh 901 TAAATGGTCTTCAAAAAGAGCAGAGAAGGAGTCCAGTGGAAATTTTGTATTAACATGA 960
Qy 1337 TGTGTATTTAATAATATCTCATGTGGAATAATATATGCACATTTGAGAAATACAGAGA 1396
Dh 961 TGTGTATTTAATAATATCTCATGTGGAATAATATATGCACATTTGAGAAATACAGAGA 1020
Qy 1397 TGAAGACAATGACATGGAAGTCCAGATATATGATGTTGAGATGATGAGTGGAGAAAC 1456
Dh 1021 TGAAGACAATGACATGGAAGTCCAGATATATGATGTTGAGATGATGAGTGGAGAAAC 1080
Qy 1457 TGTTTTCAGAGATGAGAAGAAAAATACCTTTGGAACATTACAGTGGAACTACAGAGA 1516
Dh 1081 TGTTTTCAGAGATGAGAAGAAAAATACCTTTGGAACATTACAGTGGAACTACAGAGA 1140
Qy 1517 AACATTTAATGCAACAGTATGCTTGTGACAGCATATGATCTTCTATATGCTGTGGCA 1576
Dh 1141 AACATTTAATGCAACAGTATGCTTGTGACAGCATATGATCTTCTATATGCTGTGGCA 1200
Qy 1577 AGCAGTATGATGCAATTTTTCGAATCCATATTTGATGTGGCACTTAACTGGAAGCAC 1636
Dh 1201 AGCAGTATGATGCAATTTTTCGAATCCATATTTGATGTGGCACTTAACTGGAAGCAC 1260
Qy 1637 ATCTACATGCTTCTTACTAGAAATTAACCTGAGATGCTGATGATATGCTACTAAAGCA 1696
Dh 1261 ATCTACATGCTTCTTACTAGAAATTAACCTGAGATGCTGATGATATGCTACTAAAGCA 1320

Qy 1697 AAATGTTACTGAATTTTCTATCATAAAGATGTACAGAAAGCGGAGAACCCGATATCTTA 1756
Dh 1321 AAATGTTACTGAATTTTCTATCATAAAGATGTACAGAAAGCGGAGAACCCGATATCTTA 1380
Qy 1757 TGTGTGAATGTTAGGAACCAAGATCTCTTAATAATTTATCCAGCTCAACAGGATTTCTA 1816
Dh 1381 TGTGTGAATGTTAGGAACCAAGATCTCTTAATAATTTATCCAGCTCAACAGGATTTCTA 1440
Qy 1817 TCCAGTGAATTAACATGATCCAGAAAGCGAAGAAATATTAAGTGGGGAATTAATAA 1876
Dh 1441 TCCAGTGAATTAACATGATCCAGAAAGCGAAGAAATATTAAGTGGGGAATTAATAA 1500
Qy 1877 AGACCTATCTTGTATTTCTAGTGTGATGATGGAATTTGGAATTTAGTCCAACCATGAAAAC 1936
Dh 1501 AGACCTATCTTGTATTTCTAGTGTGATGGAATTTGGAATTTAGTCCAACCATGAAAAC 1560
Qy 1937 AGCAAAAAGAAATTTTATGTAAGCAGGAACCTACCTAAAGGATATGTATCATCTGGAAT 1996
Dh 1561 AGCAAAAAGAAATTTTATGTAAGCAGGAACCTACCTAAAGGATATGTATCATCTGGAAT 1620
Qy 1997 TTATTTCAAGAAAGATGTTTGTACTGTCAAGTCAACCAATATGCTCAAGTCTTCCAGCCCT 2056
Dh 1621 TTATTTCAAGAAAGATGTTTGTACTGTCAAGTCAACCAATATGCTCAAGTCTTCCAGCCCT 1680
Qy 2057 GCTGCTTGCCAGACACAGAGGCAAAATATGAGAGCATCCCACTAGTACACACATGC 2116
Dh 1681 GCTGCTTGCCAGACACAGAGGCAAAATATGAGAGCATCCCACTAGTACACACATGC 1740
Qy 2117 ACAAGACATGTTCAAAATTAACAGATGCACCTGGAATTTGTTCCGGAATCACCTGT 2176
Dh 1741 ACAAGACATGTTCAAAATTAACAGATGCACCTGGAATTTGTTCCGGAATCACCTGT 1800
Qy 2177 GGAATAATCTTCCAGTTATTTCAAGCTTCAGAAACCAATTATTTGTTTGTCACTGATGG 2236
Dh 1801 GGAATAATCTTCCAGTTATTTCAAGCTTCAGAAACCAATTATTTGTTTGTCACTGATGG 1860
Qy 2237 CACTGTAAATCTCAATATTAAGAAAGCAATATTTACACACTGTTAAAGCAAAATCTTGG 2296
Dh 1861 CACTGTAAATCTCAATATTAAGAAAGCAATATTTACACACTGTTAAAGCAAAATCTTGG 1920
Qy 2297 TTTCAATTTACGTCAGTGGTTAAATCTAAAGAAATGTCAGTGGGAGAGAAATCTTGGC 2356
Dh 1921 TTTCAATTTACGTCAGTGGTTAAATCTAAAGAAATGTCAGTGGGAGAGAAATCTTGGC 1960
Qy 2357 GGCATATTTGATCTTGCCTGCCCTTCCCTCTTCTTGTGTTGTTGTAATCTGATTCAGG 2416
Dh 1981 GGCATATTTGATCTTGCCTGCCCTTCCCTCTTCTTGTGTTGTTGTAATCTGATTCAGG 2040
Qy 2417 TGGCCAAAGTATTTGATTTCTTTCAGACAGGCTATATTTGAAGAAACCTTGTATGTG 2476
Dh 2041 TGGCCAAAGTATTTGATTTCTTTCAGACAGGCTATATTTGAAGAAACCTTGTATGTG 2100
Qy 2477 GCTGAAGAAATTTAGACAGAGCTAGAAATCATATCAAAATTTTACTGCTCAAGATG 2536
Dh 2101 GCTGAAGAAATTTAGACAGAGCTAGAAATCATATCAAAATTTTACTGCTCAAGATG 2160
Qy 2537 GAAACCTCTCTTCCACCTATATGATTTTCTAATGATATGATGATCCCAACATCTCAAGC 2596
Dh 2161 GAAACCTCTCTTCCACCTATATGATTTTCTAATGATATGATGATCCCAACATCTCAAGC 2220
Qy 2597 TGGCAGTATGAAATTTCCCAAGTGTATGAAAGAAACAGATGTGAGAGAAATGATTAAGA 2656
Dh 2221 TGGCAGTATGAAATTTCCCAAGTGTATGAAAGAAACAGATGTGAGAGAAATGATTAAGA 2280
Qy 2657 ACAACATGAGATTAATTCGGCAGTCAGAAAAGAACCATTTGAATCTGGAATTAAGA 2716
Dh 2281 ACAACATGAGATTAATTCGGCAGTCAGAAAAGAACCATTTGAATCTGGAATTAAGA 2340
Qy 2717 TTGGAATAGAAATTAATTTGTTTAAAGAGCAAGAAAATCAATTTAGAGCTGATAAAGATT 2776
Dh 2341 TTGGAATAGAAATTAATTTGTTTAAAGAGCAAGAAAATCAATTTAGAGCTGATAAAGATT 2400

```

OY 2777 AGGATGCTCAAAAGTGAAGTAATTTTATAGGGCTGTGCTTCCAAATTTTGGCATG 2836
    |||
Db 2401 AGGATGCTCAAAAGTGAAGTAATTTTATAGGGCTGTGCTTCCAAATTTTGGCATG 2460
OY 2837 ATAGACTAATTTATTTCTTAAGAAATATATTAATCATTTTCAGTTTGGAGCTAGT 2896
    |||
Db 2461 ATAGACTAATTTATTTCTTAAGAAATATATTAATCATTTTCAGTTTGGAGCTAGT 2520
OY 2897 GCCATCCAAATGAATTAATTAATAGTACATATTTTATTAATTTTCTGTAACTAC 2956
    |||
Db 2521 GCCATCCAAATGAATTAATTAATAGTACATATTTTATTAATTTTCTGTAACTAC 2580
OY 2957 ATTAACCAAGTAAAGTGAAGGAGGCAAAATATTTTGATATTAATCTTTTCCAGCATAG 3016
    |||
Db 2581 ATTAACCAAGTAAAGTGAAGGAGGCAAAATATTTTGATATTAATCTTTTCCAGCATAG 2640
OY 3017 TATNCCCAAAATAGGAATTAATTAATTAATAGATATTTTATTAATTTTCTTTTGA 3076
    |||
Db 2641 TATNCCCAAAATAGGAATTAATTAATTAATAGATATTTTATTAATTTTCTTTTGA 2700
OY 3077 CCAAGTCTTCTAATAGCAGTACATATTTTATPACTACTGATCTTCTTCTCCGAGTAGC 3136
    |||
Db 2701 CCAAGTCTTCTAATAGCAGTACATATTTTATPACTACTGATCTTCTTCTCCGAGTAGC 2760
OY 3137 CATATTTCAAGTGTTCATTTGCCACATGTGCCCTGTGACTACTGTATTTGACAGTTCAGTA 3196
    |||
Db 2761 CATATTTCAAGTGTTCATTTGCCACATGTGCCCTGTGACTACTGTATTTGACAGTTCAGTA 2820
OY 3197 CTAGACAAAACCTAGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3256
    |||
Db 2821 CTAGACAAAACCTAGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2880
OY 3257 TAAACTTAATCAGTAACTGACAGTCCAGTCCATTCATGAGTGTGACAGTATTAATTTGTTT 3316
    |||
Db 2881 TAAACTTAATCAGTAACTGACAGTCCAGTCCATTCATGAGTGTGACAGTATTAATTTGTTT 2940
OY 3317 TATTGAGTCAATGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3376
    |||
Db 2941 TATTGAGTCAATGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3000
OY 3377 CGGGAGACATTTTGTATGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 3436
    |||
Db 3001 CGGGAGACATTTTGTATGTGATGATGATGATGATGATGATGATGATGATGATGATG 3060
OY 3437 TGGATCCCTGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3496
    |||
Db 3061 TGGATCCCTGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3120
OY 3497 AAATGCTAGTGTGCTGCAAGGCTGAGTAACCTGTGTAAAGTAACCTGTGCGACAGTAG 3556
    |||
Db 3121 AAATGCTAGTGTGCTGCAAGGCTGAGTAACCTGTGTAAAGTAACCTGTGCGACAGTAG 3180
OY 3557 GTTTCAGCAATTTCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3616
    |||
Db 3181 GTTTCAGCAATTTCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3240
OY 3617 GATATGTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3676
    |||
Db 3241 GATATGTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3300
OY 3677 ATTTTCAGAAAGTGAATGGAATTAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3736
    |||
Db 3301 ATTTTCAGAAAGTGAATGGAATTAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360
OY 3737 GTGAGTCTTTAAACCTCTCTTATTTTCTACAGATTAATGCTTAATTTTACAGATTGAACA 3796
    |||
Db 3361 GTGAGTCTTTAAACCTCTCTTATTTTCTACAGATTAATGCTTAATTTTACAGATTGAACA 3420
OY 3797 GGGATTTCAGCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3856
    |||
Db 3421 GGGATTTCAGCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3480
OY 3857 TGAATCTACCCCTTGGCAGGCTTCAGACAAATCAGTTGATCTTCCCTGAGGCCACAGCCCTC 3916

```

```

Db 3481 TGAATCTACCCCTTGGCAGGCTTCAGACAAATCAGTTGATCTCCCTGAGGCCACAGCCCTC 3540
OY 3917 ATTCTGTGAGGAGGAGAAAGATTAGCAAGAGATTAAATTTTATTCACAAATCAGTAGGT 3976
    |||
Db 3541 ATTCTGTGAGGAGGAGAAAGATTAGCAAGAGATTAAATTTTATTCACAAATCAGTAGGT 3600
OY 3977 GTTAGACATGATCTGTTGTGAGCAGTGTGTTGTCTCATATTTTCTGCTGTGACATTTTGGAG 4036
    |||
Db 3601 GTTAGACATGATCTGTTGTGAGCAGTGTGTTGTCTCATATTTTCTGCTGTGACATTTTGGAG 3660
OY 4037 ACATTTTGTGAGCAATTTCTATTTTGTGCTGTACGTATATTTTCTTTTAAATTCAGT 4096
    |||
Db 3661 ACATTTTGTGAGCAATTTCTATTTTGTGCTGTACGTATATTTTCTTTTAAATTCAGT 3720
OY 4097 GATATCTGTGCTTTAAATTTCTTACATATGATGTTTGGCTGATTCACATGATTTTATA 4156
    |||
Db 3721 GATATCTGTGCTTTAAATTTCTTACATATGATGTTTGGCTGATTCACATGATTTTATA 3780
OY 4157 ACTGAATTTTAAAGATATCAAGCTAAACAGTAAACAGTAAAGTCAATTTTCTTTATTAACA 4216
    |||
Db 3781 ACTGAATTTTAAAGATATCAAGCTAAACAGTAAACAGTAAAGTCAATTTTCTTTATTAACA 3840
OY 4217 TAGACCCGTTGCTACTCTCAGACACCCTCTCTCAATTTTCTTCTGTAGCATGTGATGC 4276
    |||
Db 3841 TAGACCCGTTGCTACTCTCAGACACCCTCTCTCAATTTTCTTCTGTAGCATGTGATGC 3900
OY 4277 CTGATTAATCAATTTTCTATTTGCTTTTATTTCTTAATATAGGACAAATAGAGTCACTC 4336
    |||
Db 3901 CTGATTAATCAATTTTCTATTTGCTTTTATTTCTTAATATAGGACAAATAGAGTCACTC 3960
OY 4337 TAAATATAGTGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAG 4396
    |||
Db 3961 TAAATATAGTGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAG 4020
OY 4397 TACACGACATAGAAACATGAATTTGCTGATTTGATACCTTTGTACAGCAATTTTGACA 4456
    |||
Db 4021 TACACGACATAGAAACATGAATTTGCTGATTTGATACCTTTGTACAGCAATTTTGACA 4080
OY 4457 GTCATTAATGTTGTGATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 4511
    |||
Db 4081 GTCATTAATGTTGTGATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 4135

RESULT 2
US-10-144-771-15372
; Sequence 15372, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: C1001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 15372
; LENGTH: 3756
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-15372

Query Match 20.6%; Score 934.6; DB 8; Length 3756;
Best Local Similarity 63.7%; Pred. No. 2,5e-186;
Matches 1859; Conservative 0; Mismatches 614; Indels 446; Gaps 11;

```

[illegible][illegible]

```

Qy 2678 AGTCAGAAAGAACCGATTGAAACTCTGGAATTAAGCATTTGGAATAGATTTGGT 2737
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2396 AGCCAGAGAGAGCGCGGTGGAATGCTGAGATTAAGCCCTGGAACACGCGCAACTGGCC 2455
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2738 TAAAGACAGAGAAATCATTTAGACGTGATTAAGAGTTAGGATGCTCAAAAGTGAAT 2797
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2456 CAAAGACACAGAGAGCCCTTTTACCACAGACAAGAGGATTATGA----- 2498
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2798 ATTTATAGAGGCTGTGGTTTCCAAAATTTTGGCATGATGACTTAATTTATTCCTT 2857
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2499 ----CGGGGCGCTGTGGGAGCGCTCTGTGTCACAGAGAAACCGAGATCAAACTGTA 2554
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2858 AAAGAAATATTAATTAATGATTTGAGTTGCGAGTTCAGTCCATCCAAATGATATATAT 2917
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2555 AAGGACAGAGACTTAATTAATGCA--TCTTAGACAGAGCCATTAAGAGGCTGTGAT 2612
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2918 ATAACTACATATTTTATTTAAATTTCTAGTACTATTAACAAAGTAAAGTGAG 2977
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2613 ACAATGAC-----ATGTAATTTCTTACTAGTACTCTTAATAAACCGTAACCGAG 2665
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2978 CAGGCAAAATATTTTGTATTTACTTTTCACCCAGTATACCCAAATAGCGAATA 3037
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2666 CAG-----ATAAACCCAGACAGAGCTTAAT 2690
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3038 TAGAATATTATTAATGATATTTTACATCCCTTTTGTACCAAGTCTTAAATGAGTA 3097
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2691 ACTAGCAAAATGTTGACAT-----TTACTTTTAATGTAAGTCTTCAAAATGACATG 2743
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3098 CATATTTTATCTTACTGATTTCTTACTCCAGTAGCAGATATTTCAAGTTCATTC 3157
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2744 C-ATTTCACACTTAATGCTTCTCTCACTAAGAGATCCGTTTTTGAAGGCTTATTG 2801
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3158 CACATGTGCGCTGTACTACTGTATGACAGCTTGAATGACATTAACAAATGACATAT 3217
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2802 CATAGTGTGATACATGACAGTGTATAGGCGAGTTCATATCAGACAAACACAGTGT 2861
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 3218 TAACCTAGTTCAGCATGATTTCTATTGATTAAT 3256
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2862 TGACCTGGGTCTAGGCATCTGTGCATTTGAATCAAT 2900
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 3 US-09-912-293-92043

```

: Sequence 92043, Application US/09912293
: GENERAL INFORMATION:
: APPLICANT: Rosen, et. al.
: TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 100
: FILE REFERENCE: PO-100
: CURRENT APPLICATION NUMBER: US/09/912,293
: PRIOR APPLICATION NUMBER: 08/103,744
: PRIOR FILING DATE: 1993-08-09
: PRIOR APPLICATION NUMBER: 09/249,651
: PRIOR FILING DATE: 1999-02-12
: PRIOR APPLICATION NUMBER: 08/104,507
: PRIOR FILING DATE: 1993-08-09
: PRIOR APPLICATION NUMBER: 08/196,363
: PRIOR FILING DATE: 1994-02-15
: PRIOR APPLICATION NUMBER: 09/859,490
: PRIOR FILING DATE: 2001-05-18
: PRIOR APPLICATION NUMBER: 08/196,362
: PRIOR FILING DATE: 1994-02-15
: PRIOR APPLICATION NUMBER: 08/221,623
: PRIOR FILING DATE: 1994-03-31
: PRIOR APPLICATION NUMBER: 08/220,691
: PRIOR FILING DATE: 1994-03-31
: PRIOR APPLICATION NUMBER: 09/741,830
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 09/813,155
: PRIOR FILING DATE: 2001-03-21
: Remaining Prior Application data removed - See file wrapper or PALM.
: NUMBER OF SEQ ID NOS: 244538

```

```

: SEQ ID NO 92043
: LENGTH: 435
: TYPE: DNA
: ORGANISM: Homo sapiens
: NAME/KEY: misc_feature
: LOCATION: (320)..(320)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (328)..(328)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (349)..(349)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (352)..(352)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (355)..(355)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (360)..(360)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (370)..(370)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (387)..(387)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (390)..(390)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (398)..(398)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (400)..(400)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (402)..(402)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (412)..(412)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (415)..(415)
: OTHER INFORMATION: n is equal to a,t,g, or c
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (433)..(433)
: OTHER INFORMATION: n is equal to a,t,g, or c
: US-09-912-293-92043
: Query Match 7.8%; Score 354.2; DB 6; Length 435;
: Best Local Similarity 94.2%; Pred. No. 9,9e-65;
: Matches 373; Conservative 0; Mismatches 22; Indels 1; Gaps 1;
Qy 2716 ATTGGAATGAGATTAATGTTTAAAGAGCAAGAAATCATTTAGCTGATTAAGAGT 2775
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 ATTCGACAGAGTAAATGTTTAAAGAGCAAGAAATCATTTAGCTGATTAAGAGT 61

```

| | | | |
|----|------|---|-------|
| QY | 2776 | TAGCATCTCTAAAGTCAGACATATTTTATATAGGGCTGTGCTTCCAAAATTTTTGGCAT | 283.5 |
| Db | 62 | TAGATGCTCAAAAGTCAGACATATTTTATATAGGGCTGTGCTTCCAAAATTTTTGGCAT | 121 |
| QY | 2836 | GATGACACTTAATTTATTTCTCTTAAAGATATATTTTAATCATTTTCAAGTTTGCAGACTAG | 289.5 |
| Db | 122 | GATGACACTTAATTTATTTCTCTTAAAGATATATTTTAATCATTTTCAAGTTTGCAGACTAG | 181 |
| QY | 2896 | TGCCATCCAAATAGATATTTATATATATTAAGTCACATATTTATTTAAATTTTCTAGTAACTA | 295.5 |
| Db | 182 | TGCCATCCAAATAGATATTTATATATATTAAGTCACATATTTATTTAAATTTTCTAGTAACTA | 241 |
| QY | 2956 | CATTAAACCAAGTAAAGTGAAGGAGGCGCAAAATATTTGATATATCTTTCACCCAGCTA | 301.5 |
| Db | 242 | CATTAAACCAAGTAAAGTGAAGGAGGCGCAAAATATTTGATATATCTTTCACCCAGCTA | 301 |
| QY | 3016 | GTATATCCCAAAATAGCGAAATATAGAATTA-TTATGAGATATTTTACATCTTTTGG | 307.4 |
| Db | 302 | GTATATCCCAAAATAGCGGATATATAGNATTTATTAATGCGGATTTTNCANCNNTTNG | 361 |
| QY | 3075 | TACCAAGCTTTCTAAATGAAGTCATATTTTATCT 3110 | |
| Db | 362 | GACCCAGTINTCTTAAATGCCGTACAAATNTTTTACT 397 | |

```

RESULT 4
US-09-513-999C-32911
Sequence 32911, Application US/09513999C
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J. B.
APPLICANT: Duclert, A. Y.
APPLICANT: Giordano, J. Y.
TITLE OR INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
FILE REFERENCE: 59.052.REG
CURRENT FILING DATE: 2000-02-24
PRIORITY APPLICATION NUMBER: US 60/122,487
PRIORITY FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 32911
LENGTH: 355
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15
OTHER INFORMATION: s-g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16
OTHER INFORMATION: k-g or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: 70
OTHER INFORMATION: s-g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: 325
OTHER INFORMATION: r-a or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: 326
OTHER INFORMATION: n-a, g, c or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: 333
OTHER INFORMATION: w-a or t
US-09-513-999C-32911
Query Match 7.58; Score 341; DB 6; Length 355;
Best Local Similarity 98.38; Pred. No. 5.6e-62;

```

| Matches | 349; | Conservative | 5; | Mismatches | 0; | Indels | 1; | Gaps | 1 |
|---------|------|--|------|------------|----|--------|----|------|---|
| QY | 2513 | CACAAATTTTACTGCTCAAGAAATGGAACCTCTCTTCCAGCTATATGATTTTCTAAGTAT | 2572 | | | | | | |
| Db | 1 | CACAATTTTACTGSGSCAAGAAATGGAACCTCTCTTCCAGCTTATGATTTTCTAAGTAT | 60 | | | | | | |
| QY | 2573 | GATGATGCGCGCAACATCTCAACGTGGCGACTGGAAGAGTTTCCCAAGTGATATGAAGAAAC | 2632 | | | | | | |
| Db | 61 | GATGATGCGSGCAACATCTCAACGTGGCGACTGGAAGAGTTTCCCAAGTGATATGAAGAAAC | 120 | | | | | | |
| QY | 2633 | AGATGTGCGAGGAATGATTAAGAAACAAACATGAAGTAATATGCGCAGTCAGAAAAAGAAC | 2692 | | | | | | |
| Db | 121 | AGATGTGCGAGGAGATGATTAAGAAACAAACATGAAGTAATATGCGCAGTCAGAAAAAGAAC | 180 | | | | | | |
| QY | 2693 | GATTGAAACTCTGGAAGATTAAGCAATTTGGAATTAATGTGTTTAAAGAACGAGAAAA | 2752 | | | | | | |
| Db | 181 | GATTGAAACTCTGGAAGATTAAGCAATTTGGAATTAATGTGTTTAAAGAACGAGAAAA | 240 | | | | | | |
| QY | 2753 | ATCATTTTGACGTGATTAAGAGATTGGAATGCTCAAAAGTGAACATAATTTATAGGGCTGT | 2812 | | | | | | |
| Db | 241 | ATCATTTTGACGTGATTAAGAGATTGGAATGCTCAAAAGTGAACATAATTTATAGGGCTGT | 300 | | | | | | |
| QY | 2813 | GGTTTCCAAAAATTTTGTGGCATGA-TAGACTTAATTTATTTTCCCTTAAAGAAATAA | 2866 | | | | | | |
| Db | 301 | GGTTTCCAAAAATTTTGTGGCATGATGATGACATGAATTAATTTATTTTCCCTTAAAGAAATAA | 355 | | | | | | |

```

RESULT 5
; US-09-912-293-91554
; Sequence 91554, Application US/09912293
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products
; FILE REFERENCE: PO-100
; CURRENT APPLICATION NUMBER: US/09/912,293
; CURRENT FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 08/103,744
; PRIOR FILING DATE: 1993-08-09
; PRIOR APPLICATION NUMBER: 09/249,651
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: 08/104,507
; PRIOR FILING DATE: 1993-08-09
; PRIOR APPLICATION NUMBER: 08/196,363
; PRIOR FILING DATE: 1994-02-15
; PRIOR APPLICATION NUMBER: 09/859,490
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 08/196,362
; PRIOR FILING DATE: 1994-02-15
; PRIOR APPLICATION NUMBER: 08/221,623
; PRIOR FILING DATE: 1994-03-31
; PRIOR APPLICATION NUMBER: 08/220,691
; PRIOR FILING DATE: 1994-03-31
; PRIOR APPLICATION NUMBER: 09/741,830
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 09/813,155
; PRIOR FILING DATE: 2001-03-21
; Remaining Prior Application data removed - See File Wrapper or PAM.
; NUMBER OF SEQ ID NOS: 244538
; SEQ ID NO 91554
; LENGTH: 313
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (310)..(310)
; OTHER INFORMATION: n is equal to a,t,g, or c
; US-09-912-293-91554

```

| | | | | |
|-----------------------|---|-------------|--------------|------------|
| Query Match | 6.9% | Score 312 | DB 6 | Length 313 |
| Best Local Similarity | 99.7% | Pred. No. 6 | 7e-56 | |
| Matches 312 | Conservative | 0 | Mismatches 1 | Indels 0 |
| Gaps | | | | 0 |
| 1911 | GGACATATTTAGTCACCAACCATGAAACAGACAAAGAGATTTTATGTGAGACGAGAAACTTAC | 1970 | | |

Db 1 GGACATTATTAGCCAAACCATGAAACAGCAAGAAAGATTTTAGTGAGGACGGAATAC 60
Qy 1971 CTAAGAATATGTATCTACGTGGAATTTATCTGAGAAGATGTTTCTACTGTCAAC 2030
Db 61 CTAAGAATATGTATCTACGTGGAATTTATCTGAGAAGATGTTTCTACTGTCAAC 120
Qy 2031 AAATATGCTGCAAGTCTCCAGCCCTGCTGCTGCGACAGACAGAGGCAAAATAGAG 2090
Db 121 AAATATGCTGCAAGTCTCCAGCCCTGCTGCTGCGACAGACAGAGGCAAAATAGAG 180
Qy 2091 AGCATCCAGTAGCTAGACACACATGACACAGATAGTTCAATATATACAGATSCACTA 2150
Db 181 AGCATCCAGTAGCTAGACACACATGACACAGATAGTTCAATATATACAGATSCACTA 240
Qy 2151 CTGGAATAGTTCCGGAATATCTGTGGAATCTCCAGTATTTAGACTTACAGAAA 2210
Db 241 CTGGAATAGTTCCGGAATATCTGTGGAATCTCCAGTATTTAGACTTACAGAAA 300
Qy 2211 CCATTATGATTT 2223
Db 301 CCATTATGATTT 313

RESULT 6
US-09-912-293-96911
; Sequence 96911, Application US/09912293
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 100
; FILE REFERENCE: PO-100
; CURRENT APPLICATION NUMBER: US/09/912, 293
; CURRENT FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 08/103,744
; PRIOR FILING DATE: 1993-08-09
; PRIOR APPLICATION NUMBER: 09/249,651
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: 08/104,507
; PRIOR FILING DATE: 1993-08-09
; PRIOR APPLICATION NUMBER: 08/196,363
; PRIOR FILING DATE: 1994-02-15
; PRIOR APPLICATION NUMBER: 09/859,490
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 08/196,362
; PRIOR FILING DATE: 1994-02-15
; PRIOR APPLICATION NUMBER: 09/741,830
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 09/813,155
; PRIOR FILING DATE: 2001-03-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 244538
; SEQ ID NO 96911
; LENGTH: 496
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10)..(10)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (160)..(160)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (184)..(184)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature

LOCATION: (214)..(214)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (232)..(232)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (238)..(238)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (285)..(285)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (287)..(287)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (332)..(333)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (347)..(347)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (353)..(353)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (367)..(367)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (404)..(404)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (407)..(407)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (411)..(411)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (420)..(420)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (426)..(426)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (439)..(439)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (443)..(443)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (445)..(445)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (472)..(472)
; OTHER INFORMATION: n is equal to a,t,g, or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (481)..(481)

OTHER INFORMATION: n is equal to a,t,g, or c
US-09-912-293-96911

| | | | | |
|-----------------------|-----------------|--------------------|------------|-------------|
| Query Match | 5.9% | Score 265.8; | DB 6; | Length 496; |
| Best Local Similarity | 83.7%; | Pred. No. 3.6e-46; | | |
| Matches 406; | Conservative 0; | Mismatches 64; | Indels 15; | Gaps 10; |

| | | | |
|----|------|---|------|
| QY | 3521 | GTAACTCTGGTTAAAAGTAACTGGGAGACATAGTTTCCAGATTTCCGTTGTCG | 3580 |
| Db | 13 | GTACACTGGTGTAAAAGTAACCTGTGGAGACATAGTTTCCAGATTTCCGTTGTCG | 72 |
| QY | 3581 | TCACGATCATGTTTGAAAAAATTTGGCTATTAAAGATATGATATAGTGGCTTTATCC | 3640 |
| Db | 73 | TCACGATCATGTTTGAAAAAATTTGGCTATTAAAGATATGATATAGTGGCTTTATCC | 132 |
| QY | 3641 | TGATATTATACCT--GGATACAACCTTGATCCCTTTCT--AATATTTTCCAGAAAGTATGGGATA | 3696 |
| Db | 133 | TGATATTATACCTGGGATACACACTTGATNTTTTCTGAAATATTTTCCAGAAAGTATGGGATA | 192 |
| QY | 3699 | ACCCTTA--GAAGAGGACTC--AGATGATATTTATATTTTAA--GTGAGCTTTAAAACCTC | 3753 |
| Db | 193 | ACCCTAGAGAGAGACTCCCGNATATGATATTTATATTTTAAAGTGAAGCTTTAAAACCTC | 252 |
| QY | 3754 | CTCTTATTTTACAAGTTATATGCTAAATTTCCAG---TTGACAGGGATCTCAGC--AT | 3808 |
| Db | 253 | CTCTTATTTTCTACAGTTATATGTGGCTAAATTTTNGATTTGGACACGGGGTTCACCACTT | 312 |
| QY | 3809 | TCTGCCATCTCTCAT--GGAAGAAGAGGCTCCCTCATCTGAAGCGTCTCT--GAATCTAC | 3866 |
| Db | 313 | CTGCCCATCTCTCATGAGGNNAGAGAGGGCTCCATCTGNNAGGGCTCTGGAATTTAC | 372 |
| QY | 3867 | CCTTGCAAGCTTCAGAACAAATTCAGTTGATCTCCGTAGGACACAGCGGCTCATCTGTGAG | 3926 |
| Db | 373 | CCTTGCAAGCTTCAGACCAATCTAGTTGGTTTTCNCGNCGACACAGGGCCT--ATTCTTTGGG | 431 |
| QY | 3927 | GGAGGGAAGATTAAGCCAAAGAGCTTAATTTTCATTTCCAAATCACCTTAACTGCTTTAGACGA | 3986 |
| Db | 432 | GGAGGGGAAGANTNGCCCAAGGGTTAATTTTCATTTCCCAANACACTGTGGGTTTACACGA | 491 |
| QY | 3987 | TCTCT 3991 | |
| Db | 492 | TCTCT 496 | |

RESULT 7
US-10-349-781-15516
; Sequence 15516, Application US/10349781
; GENERAL INFORMATION:
; APPLICANT: Sellhame, Jeffrey J.; Delegeane, Angelo M.
; APPLICANT: Stunart, Susan G.; Stuve, Laura L.
; APPLICANT: Mullaly, Sara J.; Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF CELLS AND TISSUES OF THE FEMALE REPRODUCTIVE ;
; FILE REFERENCE: PP-1028-2 CON
; CURRENT APPLICATION NUMBER: US/10/349,781
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: 09/540,764
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 08/992,868
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/032,838
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: 08/734,050
; PRIOR FILING DATE: 1996-10-18
; PRIOR APPLICATION NUMBER: 60/006,111
; PRIOR FILING DATE: 1995-10-24
; PRIOR APPLICATION NUMBER: 08/733,814
; PRIOR FILING DATE: 1996-10-18
; PRIOR APPLICATION NUMBER: 60/005,864
; PRIOR FILING DATE: 1995-10-26
; PRIOR APPLICATION NUMBER: 08/763,920
; PRIOR FILING DATE: 1996-12-11
; PRIOR APPLICATION NUMBER: 60/008,794
; PRIOR FILING DATE: 1995-12-14

```

:
: PRIOR APPLICATION NUMBER: 08/772,783
:
: PRIOR FILING DATE: 1996-12-23
:
: Remaining Prior Application data removed - See File Wrapper or PALM.
:
: NUMBER OF SEQ ID NOS: 61458
:
: SOFTWARE: PERL Program
:
: SEQ ID NO 15516
:
: LENGTH: 257
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
: FEATURE:
:
: NAME/KEY: misc_feature
:
: OTHER INFORMATION: Incyte ID NO: hu00067190
:
: US-10-349-781-15516

```

| | | | | |
|-----------------------|--------|--|--------|--------------|
| Query Match | 5.7% | Score 257 | DB 8 | Length 257 |
| Best Local Similarity | 100.0% | Pred. No. 2.2e-44 | | |
| Matches | 257 | Conservative | 0 | Mismatches 0 |
| | | | Indels | 0 |
| | | | Gaps | 0 |
| QY | 3845 | CTGAAGCCTCTGAAATCTACCCCTTGCAAGCTTGAGACAAATCAATGATGTCCTGTAG | 3904 | |
| Db | 1 | CTGAAGCCTCTCTGAAATCTACCCCTTGCAAGCTTGAGACAAATCAATGATGTCCTGTAG | 60 | |
| QY | 3905 | CCACACGGCCTCATCTCTGAGGAGGAGGAAAGATTAGCCAAAGAGTAAATTTCATCCCA | 3964 | |
| Db | 61 | CCACACGGCCTCATCTCTGAGGAGGAGGAAAGATTAGCCAAAGAGTAAATTTCATCCCA | 120 | |
| QY | 3965 | AATCACTTAGCTGTTAGACTGATCTGTTTGTGACAGTTGTTGTCTCATTTTGGCTCTGT | 4024 | |
| Db | 121 | AATCACTTAGCTGTTAGACTGATCTGTTTGTGACAGTTGTTGTCTCATTTTGGCTCTGT | 180 | |
| QY | 4025 | GCATTTTTTGAGACATTTGTTGAGAAATATTCATATTGGTGCTCACTAGTATTTTCTTTT | 4084 | |
| Db | 181 | GCATTTTTTGAGACATTTGTTGAGAAATATTCATATTGGTGCTCACTAGTATTTTCTTTT | 240 | |
| QY | 4085 | TAATATCTCACTTGATAT | 4101 | |
| Db | 241 | TAATATCTCACTTGATAT | 257 | |

```

RESULT 8
US-10-349-781-36747
: Sequence 36747, Application US/10349781
: GENERAL INFORMATION:
: APPLICANT: Sellhameer, Jeffrey J.; Delegeane, Angelo M.
: APPLICANT: Stuart, Susan G.; Stuve, Laura L.
: APPLICANT: Mullaly, Sara J.; Naughton, Rebecca E.
: TITLE OF INVENTION: POLYNUCLEOTIDES OF CELLS AND TISSUES OF THE FEMALE REPRODUCTIVE
: FILE REFERENCE: PD-1028-2-CON
: CURRENT APPLICATION NUMBER: US/10/349,781
: CURRENT FILING DATE: 2003-01-21
: PRIOR APPLICATION NUMBER: 09/540,764
: PRIOR FILING DATE: 2000-03-30
: PRIOR APPLICATION NUMBER: 08/992,868
: PRIOR FILING DATE: 1997-12-11
: PRIOR APPLICATION NUMBER: 60/032,838
: PRIOR FILING DATE: 1996-12-13
: PRIOR APPLICATION NUMBER: 08/734,050
: PRIOR FILING DATE: 1996-10-18
: PRIOR APPLICATION NUMBER: 60/006,111
: PRIOR FILING DATE: 1995-10-24
: PRIOR APPLICATION NUMBER: 08/733,814
: PRIOR FILING DATE: 1996-10-18
: PRIOR APPLICATION NUMBER: 60/005,864
: PRIOR FILING DATE: 1995-10-26
: PRIOR APPLICATION NUMBER: 08/763,920
: PRIOR FILING DATE: 1996-12-11
: PRIOR APPLICATION NUMBER: 60/008,794
: PRIOR FILING DATE: 1995-12-14
: PRIOR APPLICATION NUMBER: 08/772,783
: PRIOR FILING DATE: 1996-12-23
: Remaining Prior Application data removed - See File Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 61458
: SOFTWARE: PERL Program

```

;; SEQ ID NO 36747
;; LENGTH: 245
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; OTHER INFORMATION: Incyte ID No: hu00936261
US-10-349-781-36747

Query Match 5.4%; Score 245; DB 8; Length 245;
Best Local Similarity 100.0%; Pred. No. 7.3e-42;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3845 CTGAAGCTCTCTGAAATCTACCTTGCAGCTCAGACAAATCAGTTGATCTCCCTGAG 3904
DB 1 CTGAAGCTCTCTGAAATCTACCTTGCAGCTCAGACAAATCAGTTGATCTCCCTGAG 60
QY 3905 CCACACGGCCCTCATCTCTGTGAGGGAGGAGAAAGATTAGCCAAAGATTATTTTCATTCCA 3964
DB 61 CCACACGGCCCTCATCTCTGTGAGGGAGGAGAAAGATTAGCCAAAGATTATTTTCATTCCA 120
QY 3965 AATCACTAGCTGTAGACTGATCTGTTGTAGACAGTTGTTGCTCATTTTGGCTGT 4024
DB 121 AATCACTAGCTGTAGACTGATCTGTTGTAGACAGTTGTTGCTCATTTTGGCTGT 180
QY 4025 GCATTTTGTGAGACATTTGTGAGAAATATCTATTTGGTCTCTACTGATTTTCTTTT 4084
DB 181 GCATTTTGTGAGACATTTGTGAGAAATATCTATTTGGTCTCTACTGATTTTCTTTT 240
QY 4085 TAATA 4089
DB 241 TAATA 245

RESULT 9

US-10-349-781-24291
;; Sequence 24291, Application US/10349781
;; GENERAL INFORMATION:
;; APPLICANT: Seilhamer, Jeffrey J.; Deleage, Angelo M.
;; APPLICANT: Stuart, Susan G.; Stuve, Laura L.
;; APPLICANT: Mulahy, Sara J.; Naughton, Rebecca E.
;; TITLE OF INVENTION: POLYNUCLEOTIDES OF CELLS AND TISSUES OF THE FEMALE REPRODUCTIVE S
;; FILE REFERENCE: PD-1028-2 CON
;; CURRENT APPLICATION NUMBER: US/10/349,781
;; PRIOR APPLICATION NUMBER: 2003-01-21
;; PRIOR FILING DATE: 2000-03-30
;; PRIOR APPLICATION NUMBER: 08/992,868
;; PRIOR FILING DATE: 1997-12-11
;; PRIOR APPLICATION NUMBER: 60/032,838
;; PRIOR FILING DATE: 1996-12-13
;; PRIOR APPLICATION NUMBER: 08/734,050
;; PRIOR FILING DATE: 1996-10-18
;; PRIOR APPLICATION NUMBER: 60/006,111
;; PRIOR FILING DATE: 1995-10-24
;; PRIOR APPLICATION NUMBER: 08/733,814
;; PRIOR FILING DATE: 1996-10-18
;; PRIOR APPLICATION NUMBER: 60/005,864
;; PRIOR FILING DATE: 1995-10-26
;; PRIOR APPLICATION NUMBER: 08/763,920
;; PRIOR FILING DATE: 1996-12-11
;; PRIOR APPLICATION NUMBER: 60/008,794
;; PRIOR FILING DATE: 1995-12-14
;; PRIOR APPLICATION NUMBER: 08/772,783
;; PRIOR FILING DATE: 1996-12-23
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 61458
;; SOFTWARE: PERL Program
;; SEQ ID NO 24291
;; LENGTH: 228
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:

;; NAME/KEY: misc_feature
;; OTHER INFORMATION: Incyte ID No: hu00693953
US-10-349-781-24291

Query Match 5.0%; Score 228; DB 8; Length 228;
Best Local Similarity 100.0%; Pred. No. 2.6e-38;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3667 CTTTCTATATTTTCAGAAAGTATGGATACCTTGAAGAGACCTCAGATGATNTT 3726
DB 1 CTTTCTATATTTTCAGAAAGTATGGATACCTTGAAGAGACCTCAGATGATNTT 60
QY 3727 TATATTTTAACTGACTCTTAAACCTCCTTATTTCTCAAGTATATGCTTAATTTTC 3786
DB 61 TATATTTTAACTGACTCTTAAACCTCCTTATTTCTCAAGTATATGCTTAATTTTC 120
QY 3787 AGATTGAACAGGAGATTGACATTTGCCATCTCTCATGAGAAAGAGAGGCTCCCTCATCT 3846
DB 121 AGATTGAACAGGAGATTGACATTTGCCATCTCTCATGAGAAAGAGAGGCTCCCTCATCT 180
QY 3847 GAAGGCTCTGAAATCTACCTTGCAGGCTTCAGACAAATCAGTTGA 3894
DB 181 GAAGGCTCTGAAATCTACCTTGCAGGCTTCAGACAAATCAGTTGA 228

RESULT 10

US-09-912-293-18131
;; Sequence 18131, Application US/09912293
;; GENERAL INFORMATION:
;; APPLICANT: Rosen, et. al.
;; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 100
;; FILE REFERENCE: PO-100
;; CURRENT APPLICATION NUMBER: US/09/912,293
;; PRIOR FILING DATE: 2001-07-26
;; PRIOR APPLICATION NUMBER: 08/103,744
;; PRIOR FILING DATE: 1993-08-09
;; PRIOR APPLICATION NUMBER: 09/249,651
;; PRIOR FILING DATE: 1999-02-12
;; PRIOR APPLICATION NUMBER: 08/104,507
;; PRIOR FILING DATE: 1993-08-09
;; PRIOR APPLICATION NUMBER: 08/196,363
;; PRIOR FILING DATE: 1994-02-15
;; PRIOR APPLICATION NUMBER: 09/859,490
;; PRIOR FILING DATE: 2001-05-18
;; PRIOR APPLICATION NUMBER: 08/196,362
;; PRIOR FILING DATE: 1994-02-15
;; PRIOR APPLICATION NUMBER: 08/221,623
;; PRIOR FILING DATE: 1994-03-31
;; PRIOR APPLICATION NUMBER: 08/220,691
;; PRIOR FILING DATE: 1994-03-31
;; PRIOR APPLICATION NUMBER: 09/741,830
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 09/813,155
;; PRIOR FILING DATE: 2001-03-21
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 244538
;; SEQ ID NO 18131
;; LENGTH: 375
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (172)..(172)
;; OTHER INFORMATION: n is equal to a,t,g, or c
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (201)..(201)
;; OTHER INFORMATION: n is equal to a,t,g, or c
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (223)..(223)
;; OTHER INFORMATION: n is equal to a,t,g, or c
;; FEATURE:

```

NAME/KEY: misc_feature
LOCATION: (236)..(236)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (239)..(239)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (242)..(242)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (267)..(267)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (275)..(275)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (304)..(304)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (306)..(306)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (327)..(327)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (339)..(339)
OTHER INFORMATION: n is equal to a,t,g, or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (370)..(370)
OTHER INFORMATION: n is equal to a,t,g, or c
US-09-912-293-18131

```

```

Query Match
Best Local Similarity 92.4%; Score 188.6; DB 6; Length 375;
Matches 218; Conservative 0; Mismatches 14; Indels 4; Gaps 2;

```

```

QY 777 GCTCTGAAAGAAACCAATATATATCTCATATATGTAAGAGCCATTGGAATACAGAG 836
      |||
DB 8 GCTCTGAAAGAAACCAATATATATCTCATATATGTAAGAGCCATTGGAATACAGAG 67
QY 837 CACAGAGCAGTATGAGACCGCTTTTGTATGAGGACTACATACCAATTTGCTTAAC 896
      |||
DB 68 CACAGAGCAGTATGAGACCGCTTTTGTATGAGGACTACATACCAATTTGCTTAAC 127
QY 897 ACAGAAATTCGCTTTGGAAGATTTGGCTTGAGAGATGTG--AATATGCACATCTCT 954
      |||
DB 128 ACAGAAATTCGCTTTGGAAGATTTGGCTTGAGAGATGTG--AATATGCACATCTCT 187
QY 955 ACTTTTTCATGTAAACTAGTCTTGG--ACTTGACCAGACATATGAGAAGACAC 1008
      |||
DB 188 ACTTTTTCATGTAAACTAGTCTTGG--ACTTGACCAGACATATGAGAAGACAC 243

```

```

RESULT 11
US-09-539-806B-30366
; Sequence 30366, Application US/09539806B
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Deleageane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF EXOCRINE GLAND TISSUE

```

```

FILE REFERENCE: PD-1027 CIP
CURRENT APPLICATION NUMBER: US/09/539,806B
CURRENT FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: 08/706,766
PRIOR FILING DATE: September 27, 1996
PRIOR APPLICATION NUMBER: 60/004,676
PRIOR FILING DATE: October 2, 1995
PRIOR APPLICATION NUMBER: 08/749,515
PRIOR FILING DATE: November 15, 1996
PRIOR APPLICATION NUMBER: 60/006,810
PRIOR FILING DATE: November 15, 1995
PRIOR APPLICATION NUMBER: 08/822,285
PRIOR FILING DATE: March 20, 1997
PRIOR APPLICATION NUMBER: 60/013,696
PRIOR FILING DATE: March 20, 1996
PRIOR APPLICATION NUMBER: 08/951,197
PRIOR FILING DATE: October 1, 1997
PRIOR APPLICATION NUMBER: 60/027,249
PRIOR FILING DATE: October 1, 1996
PRIOR APPLICATION NUMBER: 08/826,438
PRIOR FILING DATE: March 20, 1997
PRIOR APPLICATION NUMBER: 60/016,145
PRIOR FILING DATE: April 18, 1996
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 48372
SOFTWARE: PERL Program
SEQ ID NO 30366
LENGTH: 181
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: hu00862945
US-09-539-806B-30366

```

```

Query Match
Best Local Similarity 100.0%; Score 181; DB 5; Length 181;
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2756 ATTAGACGTGATTAAGAGTTAGATGCTCAAAAGTGAACATATTTATAGGCGCTGTGT 2815
      |||
DB 1 ATTAGACGTGATTAAGAGTTAGATGCTCAAAAGTGAACATATTTATAGGCGCTGTGT 60
QY 2816 TTCCAAATTTTGGCATGATAGCTTAATTTATTTCTTAAGACATATTTAAATC 2875
      |||
DB 61 TTCCAAATTTTGGCATGATAGCTTAATTTATTTCTTAAGACATATTTAAATC 120
QY 2876 ATTCAAGTTTGAGAGCTGAGTCGATCCATCAATGAATTAATATAGTACATATTTAT 2935
      |||
DB 121 ATTCAAGTTTGAGAGCTGAGTCGATCCATCAATGAATTAATATAGTACATATTTAT 180
QY 2936 T 2936
DB 181 T 181

```

```

RESULT 12
US-10-349-781-1729
; Sequence 1729, Application US/10349781
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.; Deleageane, Angelo M.
; APPLICANT: Stuart, Susan G.; Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.; Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF CELLS AND TISSUES OF THE FEMALE REPRODUCTIV
; FILE REFERENCE: PD-1028-2 CON
; CURRENT APPLICATION NUMBER: US/10/349,781
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: 09/540,764
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 08/992,868
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/032,838
; PRIOR FILING DATE: 1996-12-13

```

```
; PRIOR APPLICATION NUMBER: 08/734,050
; PRIOR FILING DATE: 1996-10-18
; PRIOR APPLICATION NUMBER: 60/006,111
; PRIOR FILING DATE: 1995-10-24
; PRIOR APPLICATION NUMBER: 08/733,814
; PRIOR FILING DATE: 1996-10-18
; PRIOR APPLICATION NUMBER: 60/005,864
; PRIOR FILING DATE: 1995-10-26
; PRIOR APPLICATION NUMBER: 08/763,920
; PRIOR FILING DATE: 1996-12-11
; PRIOR APPLICATION NUMBER: 60/008,794
; PRIOR FILING DATE: 1995-12-14
; PRIOR APPLICATION NUMBER: 08/772,783
; PRIOR FILING DATE: 1996-12-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 61458
; SOFTWARE: PERL Program
; SEQ ID NO 1729
; LENGTH: 191
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu00876731
US-10-349-781-1729

Query Match
Best Local Similarity 4.0%; Score 179; DB 8; Length 191;
Matches 190; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 655 GAGAAATCCCTACTGACACCTTTGTTGATGTAATGCCATGTGCGCCATGTTCTCTTTG 714
DB 1 GAGAAATCCCTACTGACACCTTTGTTGATGTAATGCCATGTGCGCCATGTTCTCTTTG 60

QY 715 CTCTCTTTTGTAGTGAAGTGAATATATATACCAACCTGGAAGACCTTGAGATAGAA 774
DB 61 CTCTCTTTTGTAGTGAAGTGAATATATATACCAACCTGGAAGACCTTGAGATAGAA 120

QY 775 ATGCTCGAAGAAAGAAATATATATCTCATATGTAAGCCCTTGGAATA-CCA 833
DB 121 ATGCTCGAAGAAAGAAATATATATCTCATATGTAAGCCCTTGGAATA-CCA 180

QY 834 GAGCAGACAGC 844
DB 181 GAGCAGACAGC 191

RESULT 13
US-10-349-781-23200
; Sequence 23200, Application US/10349781
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.; Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.; Stuve, Laura L.
; APPLICANT: Mullaly, Sara J.; Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES OF CELLS AND TISSUES OF THE FEMALE REPRODUCTIVE S
; FILE REFERENCE: PD-1028-2 CON
; CURRENT APPLICATION NUMBER: US/10/349,781
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: 09/540,764
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 08/992,868
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/032,838
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: 08/734,050
; PRIOR FILING DATE: 1996-10-18
; PRIOR APPLICATION NUMBER: 60/006,111
; PRIOR FILING DATE: 1995-10-24
; PRIOR APPLICATION NUMBER: 08/733,814
; PRIOR FILING DATE: 1996-10-18
; PRIOR APPLICATION NUMBER: 60/005,864
; PRIOR FILING DATE: 1995-10-26
; PRIOR APPLICATION NUMBER: 08/763,920
```

```
; PRIOR FILING DATE: 1996-12-11
; PRIOR APPLICATION NUMBER: 60/008,794
; PRIOR FILING DATE: 1995-12-14
; PRIOR APPLICATION NUMBER: 08/772,783
; PRIOR FILING DATE: 1996-12-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 61458
; SOFTWARE: PERL Program
; SEQ ID NO 23200
; LENGTH: 176
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu00313647
; NAME/KEY: unsure
; LOCATION: 137
; OTHER INFORMATION: a, t, c, g, or other
US-10-349-781-23200

Query Match
Best Local Similarity 3.8%; Score 171; DB 8; Length 176;
Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3530 TGTAAAGTAACCGTGGCAGACTAGAGTTCCAGAAATTCGGTGTGCTACAGTATC 3589
DB 1 TGTAAAGTAACCGTGGCAGACTAGAGTTCCAGAAATTCGGTGTGCTACAGTATC 60

QY 3590 ATGTTTGAAGAAATTTTGGCTATTAAGATATGTAATGATGCTTATTCGTGATATTA 3649
DB 61 ATGTTTGAAGAAATTTTGGCTATTAAGATATGTAATGATGCTTATTCGTGATATTA 120

QY 3650 CCTGGATCAACCTGATCTTTTCTAATATTTTTCAGAAAGTATGGATGATAC 3701
DB 121 CCTGGATCAACCTGATCTTTTCTAATATTTTTCAGAAAGTATGGATGATAC 172

RESULT 14
US-09-947-911-240/c
; Sequence 240, Application US/09947911
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPs) LOCATED ON
; FILE REFERENCE: CI001300
; CURRENT APPLICATION NUMBER: US/09/947,911
; CURRENT FILING DATE: 2001-09-07
; NUMBER OF SEQ ID NOS: 368
; SEQ ID NO 240
; LENGTH: 1043426
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)-(1043426)
; OTHER INFORMATION: n = A,T,C or G
US-09-947-911-240

Query Match
Best Local Similarity 2.4%; Score 108.4; DB 5; Length 1043426;
Matches 253; Conservative 0; Mismatches 161; Indels 8; Gaps 4;

QY 2799 TTTTATAGGCTGTGGTTCCAAAATTTTGGCATATGACTTAATTTATTTCCCTTA 2858
DB 847510 TTTTCCAAAATGATTTTATCATTTTATCAGACATTTTAAAGTAAGTTTAAATTT 847451

QY 2859 AAGATATATTAATATCATTTTCAAGTTTGCAGACAGTGCATCCAAATGAAATATATA 2918
DB 847450 TAAAAACATGTATAGCTTTTAAATGATTTGAAATGCTATTTTAAATGAAATATATA 847391

QY 2919 TAAATCATATTTTATTAATTAATTTTCTAGTAATCAATTAACAAAGTAAAGTGAGC 2978
DB 847391 TAAATCATATTTTATTAATTAATTTTCTAGTAATCAATTAACAAAGTAAAGTGAGC 2978
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2003, 04:56:46 ; Search time 6593 Seconds
(without alignments)
1117.970 Million cell updates/sec

Title: US-09-847-046-1
Perfect score: 4526
Sequence: 1 gccgagtcgctgctgctgc.....cctcaaaaaaaaaaaaaa 4526

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estnu: *
4: em_estcov: *
5: em_estpl: *
6: em_estro: *
7: em_estro: *
8: em_hlc: *
9: gb_est1: *
10: gb_est2: *
11: gb_hlc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rtd: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 776.2 | 17.1 | 784 | 9 | AU124110 AU124110 |
| 2 | 714.4 | 15.8 | 847 | 12 | BG546766 602574136 |
| 3 | 710.2 | 15.7 | 753 | 14 | BM728814 UT-E-E01- |
| 4 | 702.8 | 15.5 | 743 | 9 | AU119721 AU119721 |
| 5 | 688.8 | 15.2 | 1266 | 11 | AK017582 Mus muscu |
| 6 | 639.2 | 14.1 | 1453 | 13 | BM552780 AGENCOURT |

| 7 | 602 | 13.3 | 605 | 10 | AV646118 |
|----|-------|------|------|----|----------|
| c | 583.6 | 12.9 | 593 | 13 | BI712516 |
| 8 | 581.2 | 12.8 | 612 | 10 | AW961708 |
| 10 | 568.2 | 12.6 | 686 | 13 | BG928131 |
| c | 567 | 12.5 | 585 | 14 | BM972493 |
| c | 563.6 | 12.5 | 564 | 9 | AL040840 |
| 13 | 562 | 12.4 | 900 | 12 | BE895218 |
| c | 556.4 | 12.3 | 695 | 12 | BF790996 |
| 14 | 552.6 | 12.2 | 556 | 10 | AW665394 |
| c | 549.8 | 12.1 | 578 | 10 | AV662311 |
| 16 | 549.8 | 12.1 | 512 | 10 | AW131594 |
| c | 497 | 11.0 | 497 | 9 | AA934775 |
| c | 489.6 | 10.8 | 751 | 12 | BG863191 |
| 18 | 485.8 | 10.7 | 488 | 13 | BI712755 |
| c | 483.4 | 10.7 | 488 | 13 | BI712755 |
| 20 | 483.4 | 10.6 | 627 | 9 | AA113856 |
| 21 | 481.6 | 10.5 | 477 | 9 | AA278257 |
| 22 | 475.4 | 10.5 | 571 | 12 | BG609542 |
| 23 | 475 | 10.4 | 470 | 9 | AI333000 |
| c | 469.6 | 10.4 | 728 | 13 | BI689134 |
| 24 | 463.8 | 10.2 | 470 | 9 | AI832055 |
| c | 461 | 10.2 | 442 | 9 | AA278675 |
| 26 | 441.6 | 9.8 | 475 | 10 | AW131649 |
| c | 439.4 | 9.7 | 439 | 9 | AI306463 |
| 28 | 427 | 9.4 | 426 | 9 | AI453818 |
| c | 425.6 | 9.4 | 425 | 9 | AI014524 |
| c | 423 | 9.3 | 604 | 9 | AA113931 |
| c | 421.2 | 9.3 | 539 | 14 | H06836 |
| 32 | 412.6 | 9.1 | 415 | 10 | AW150811 |
| c | 410.6 | 9.1 | 887 | 14 | BQ919809 |
| 34 | 408 | 9.0 | 416 | 9 | AI140821 |
| c | 407.6 | 9.0 | 807 | 13 | BI655128 |
| 37 | 407.2 | 9.0 | 1077 | 13 | BI689253 |
| c | 398.6 | 8.8 | 383 | 9 | AA773166 |
| 39 | 383 | 8.5 | 410 | 9 | AI333002 |
| c | 376.8 | 8.3 | 385 | 9 | AA737677 |
| 40 | 368.4 | 8.1 | 1177 | 14 | BQ215516 |
| c | 355.6 | 7.9 | 588 | 14 | BM941478 |
| 42 | 329.6 | 7.3 | 604 | 12 | BF458484 |
| c | 329.6 | 7.3 | 603 | 12 | BF458491 |
| 44 | 328 | 7.2 | | | |
| 45 | | | | | |

ALIGNMENTS

RESULT 1
AU124110
LOCUS AU124110 NT2RM2 Homo sapiens cDNA clone NT2RM201681 5', mRNA
DEFINITION AU124110 784 bp mRNA linear EST 01-AUG-2002
ACCESSION AU124110
VERSION AU124110.1 GI:10948826
KEYWORDS EST.
SOURCE Human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 784)
Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
Isogai,T.
HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
Y., Sugano,S., Isogai,T.)
Unpublished (2000)

JOURNAL COMMENT
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of

Virology, Institute of Medical Science, University of Tokyo, and
Heliix Research Institute.

FEATURES
Location/Qualifiers
1..784

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RM2001681"
/clone_lib="NT2RM2"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal precursor cells"

BASE COUNT 262 a 149 c 148 g 222 t 3 others
ORIGIN

Query Match 17.1%; Score 776.2; DB 9; Length 784;
Best Local Similarity 99.2%; Pred. No. 3.9e-137;

Matches 778; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1686 TGTACTAGCAAAATGTACTGTAATTCCTATCATTAAGATGTACAGAAAGCGAGAAC 1745

Db 1 TGTACTAGCAAAATGTACTGTAATTCCTATCATTAAGATGTACAGAAAGCGAGAAC 60

QY 1746 CCGATATCTTATGCTGAATGTAGAACCAAGATCTCTAAATTTATCCAGCTCAAC 1805

Db 61 CCGATATCTTATGCTGAATGTAGAACCAAGATCTCTAAATTTATCCAGCTCAAC 120

QY 1806 AGGATTTCAATTCAGTGAATATATACATGCATCCAGAAAGCAAGAAATTTTAAGTGG 1865

Db 121 AGGATTTCAATTCAGTGAATATATACATGCATCCAGAAAGCAAGAAATTTTAAGTGG 180

QY 1866 GAATTTATTAAGACCTCATCTTGTATTTCTAGTGTCTAGTATTTGGACTATTTAGTCCA 1925

Db 181 GAATTTATTAAGACCTCATCTTGTATTTCTAGTGTCTAGTATTTGGACTATTTAGTCCA 240

QY 1926 ACCATGAAACAGCAAAAGAGATTTTGTAGAACGGAAGCACTAAAGAGATATGTT 1985

Db 241 ACCATGAAACAGCAAAAGAGATTTTGTAGAACGGAAGCACTAAAGAGATATGTT 300

QY 1986 ATCACTGGAATTTATCTGAGAGAGATGTTTGTCTACTGTCAACCAAAATATGCTGCAAGT 2045

Db 301 ATCACTGGAATTTATCTGAGAGAGATGTTTGTCTACTGTCAACCAAAATATGCTGCAAGT 360

QY 2046 CTTCACGCCCTGCTGCTGGCCAGACACAGAAAGCAAAATATGAGAGCATCCCACTAGCT 2105

Db 361 CTTCACGCCCTGCTGCTGGCCAGACACAGAAAGCAAAATATGAGAGCATCCCACTAGCT 420

QY 2106 AGCAGCATGCAACAGATAGTTCAATTAATTAACAGAGATGCACTAGTGAATGTTCCG 2165

Db 421 AGCAGCATGCAACAGATAGTTCAATTAATTAACAGAGATGCACTAGTGAATGTTCCG 480

QY 2166 GAATTCACGTGGAAATCTTCCAGTATTTTCAGACTTCAGAAACCATTAATGATTTTG 2225

Db 481 GAATTCACGTGGAAATCTTCCAGTATTTTCAGACTTCAGAAACCATTAATGATTTTG 540

QY 2226 TTTCAGTGAAGCACTGTAATCTCAATATTAAGAAAGCAATATTTGACACTGTTAAAGCAG 2285

Db 541 TTTCAGTGAAGCACTGTAATCTCAATATTAAGAAAGCAATATTTGACACTGTTAAAGCAG 600

QY 2286 AAATCTTGGAATTCATTTCTCCATGCTGTAAATCTTAAGAAATCTCCAGTGGGAGAA 2345

Db 601 AAATCTTGGAATTCATTTCTCCATGCTGTAAATCTTAAGAAATCTCCAGTGGGAGAA 660

QY 2346 GGAATTCCTGGGCAATATTTGATCCTTGCCCTCCCTTCCTTGTGTTGGTGAAT 2405

Db 661 GGAATTCCTGGGCAATATTTGATCCTTGCCCTCCCTTCCTTGTGTTGGTGAAT 720

QY 2406 CTGCATTCAGGTGGCCAGATTTTGCATTTCTTCAGACCGAGCTATTAATTTGAAGAAAC 2465

Db 721 CTGCATTCAGGTGGCCAGATTTTGCATTTCTTCAGACCGAGCTATTAATTTGAAGAAAC 780

QY 2466 CTTG 2469
||||

Db 781 CTTG 784

RESULT 2
LOCUS BG546766 847 bp mRNA linear EST 04-APR-2001

DEFINITION 602574136F1 NIH_MGC_77 Homo sapiens CDNA clone IMAGE:4702136 5',

ACCESSION BG546766

VERSION BG546766.1 GI:13545431

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 847)

TITLE NIH-MGC http://imgc.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cga@bbs-r@mail.nih.gov

Tissue Procurement: CLOMTECH Laboratories, Inc.

CDNA Library Preparation: CLOMTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

plate: LCM1538 row: h column: 09

High quality sequence stop: 809.

Location/Qualifiers

1..847

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="NIH_MGC_77"

/lab_host="DH10B (TI phage-resistant)"

/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:

5' (ggcgccctcgcc); Site_2: 5' (ggcgccctcgcc); 5' and

3' adaptors were used in cloning as follows: 5' adaptor

sequence: 5'-CACGGCCATTAAGGCC-3' and 3' adaptor sequence:

5'-ATTCTAGAGCGCGAGCGCGCCAGCAGT-dT(30)-BT-3' (where B = A,

C, or G and N = A, C, G, or T). Average insert size 1.9

kb (range 0.5-4.0 kb). 12/15 colonies contained inserts

by PCR. This library was enriched for full-length clones

and was constructed by Clontech Laboratories (Palo Alto,

CA). Note: this is a NIH-MGC Library."

BASE COUNT 251 a 156 c 158 g 282 t

ORIGIN

Query Match 15.8%; Score 714.4; DB 12; Length 847;

Best Local Similarity 94.1%; Pred. No. 2e-125;

Matches 798; Conservative 0; Mismatches 41; Indels 9; Gaps 5;

QY 3027 ATAGCAAAATATGAATTAATTAATGATATTTTACATCCTTTTGTACCAAGCTTC 3086

Db 1 ATAGCAAAATATGAATTAATTAATGATATTTTACATCCTTTTGTACCAAGCTTC 60

QY 3087 TAAATGACAGACATATTTTATCTCTGATTTCTTACTTCGAGTAGCATATTTTCAA 3146

Db 61 TAAATGACAGACATATTTTATCTCTGATTTCTTACTTCGAGTAGCATATTTTCAA 120

QY 3147 GTGTCATTGGCCAGATGTGGCCGTGACTACTGTATTTGACAGTTGACTACTAGACAAA 3206

Db 121 GTGTCATTGGCCAGATGTGGCCGTGACTACTGTATTTGACAGTTGACTACTAGACAAA 180

QY 3207 ACTAGCATTAATTAATTTAGTTTCTAGCCATGATTTTCTAATTTGATTAATTAACCTTA 3266

Db 181 ACTAGCATTAATTAATTTAGTTTCTAGCCATGATTTTCTAATTTGATTAATTAACCTTA 240

QY 3267 TCACAGTTAACTCCACAGGCAATTCATGACGCGGAGAGTATATTTGTTTATTTGGAGTC 3326

Db 241 TCACAGTTAACTCCACAGGCAATTCATGACGCGGAGAGTATATTTGTTTATTTGGAGTC 300

|||||
 Db 537 TATGTTAGATAGCTTTATCCCTGATTAATACCTGATACACCTGATCTTTCTAATAT 596
 Oy 3679 TTTCAGAAAGTATGGATTAACCTGAGAGAGACTCAGATGATATTTATTTAAAGT 3738
 Db 597 TTTCAGAAAGTATGGATTAACCTGAGAGAGACTCAGATGATATTTATTTAAAGT 654
 Oy 3739 GAGCTTTAAACCTCTCTTATTTTACAACTTATATGGCTAAATTTTCAGATTTGAACAG 3798
 Db 655 GAGCTTTAAACCTCTCTTATTTTACAACTTATATGGCTAAATTTTCAGATTTGAACAG 714
 Oy 3799 GATTTCAGCATTTCTCCATCTCTCATGGAAGAGAGGCT 3837
 Db 715 GATTTCAGCATTTCTCCATCTCTCATGGAAGAGAGGCT 753

RESULT 4
 AUI19721 AUI19721 743 bp mRNA linear EST 01-AUG-2002
 LOCUS AUI19721 HEMBA1 Homo sapiens CDNA clone HEMBA1006479 5', mRNA
 DEFINITION sequence.
 ACCESSION AUI19721
 VERSION AUI19721.1 GI:10934956
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 743)
 Ota,T., Nishikawa,T., Suzuki,Y., Ishi,S., Saito,K., Kawai,Y.,
 Yamamoto,Y., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
 Isoqal,T.

TITLE HRI human CDNA project
 JOURNAL Unpublished (2000)
 COMMENT Contact: Takao Isoqal
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3975
 Fax: 81-438-52-3986
 Email: genomics@hri.co.jp

HRI human CDNA project, 5'- & 3'-end one pass sequencing: Helix
 Research Institute; CDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.

FEATURES
 Location/Qualifiers
 source 1..743
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HEMBA1006479"
 /clone_lib="HEMBA1"
 /tissue_type="whole embryo, mainly head"
 /dev_stage="embryo, 10 weeks"
 /note="Vector: pME18SFL3"

BASE COUNT 197 a 141 c 147 g 254 t 4 others
 ORIGIN

Query Match 15.5%; Score 702.8; DB 9; Length 743;
 Best Local Similarity 99.4%; Pred. No. 3.3e-123;
 Matches 715; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Oy 3361 GATATTAGCAATTTGCGGAGACATTTTGTATGATGATGATGAGGACAGTTATGACATT 3420
 Db 1 GATATTAGCAATTTGCGGAGACATTTTGTATGATGATGATGAGGACAGTTATGACATT 60
 Oy 3421 TAGTGATGAGAGCCATGATCTGCTAATAATACCTGATGGACAGCCGCCACACAA 3480
 Db 61 TAGTGATGAGAGCCATGATCTGCTAATAATACCTGATGGACAGCCGCCACACAA 120
 Oy 3481 AGAATATATCTGCGGCAAAATGATGCTGCGCAAGGCTGATTAACCTTGTTAAAGTA 3540
 Db 121 AGAATATATCTGCGGCAAAATGATGCTGCGCAAGGCTGATTAACCTTGTTAAAGTA 180

Oy 3541 ACCTGTGGAGACTAGGTTTCCAGAAATTCCTGGTTCTGCTCAGTATCATGTTGAAAA 3600
 Db 181 ACCTGTGGAGACTAGGTTTCCAGAAATTCCTGGTTCTGCTCAGTATCATGTTGAAAA 240
 Oy 3601 AATTTGGCATTAAGATATGATTA- GATGCTTATTCCTGATATTAACCTGGATTA 3659
 Db 241 AATTTGGCATTAAGATATGATTAAGATGATGATGATGATGATGATGATGATGATGAT 300
 Oy 3660 ACTTATCTTTTCTAATATTTTTCAGAAAGTATGATGATGATGATGATGATGATGATGAT 3719
 Db 301 ACTTATCTTTTCTAATATTTTTCAGAAAGTATGATGATGATGATGATGATGATGATGAT 360
 Oy 3720 TGATATTTATTTTAAAGTACTTAAACCTCTCTTATTTTTCAGAAAGTATGATGCT 3779
 Db 361 TGATATTTATTTTAAAGTACTTAAACCTCTCTTATTTTTCAGAAAGTATGATGCT 420
 Oy 3780 AATTTTCAGATTGACAGGATTCAGCATTCCTCCATCTCCATGATGAAAGAGGCTCC 3839
 Db 421 AATTTTCAGATTGACAGGATTCAGCATTCCTCCATCTCCATGATGAAAGAGGCTCC 480
 Oy 3840 CTCATCTGAAGCTCTCTGAAATCTACCTTGCACACTTCAGACAAATGATGATCTCC 3899
 Db 481 CTCATCTGAAGCTCTCTGAAATCTACCTTGCACACTTCAGACAAATGATGATCTCC 540
 Oy 3900 CTGAGCCACAGGCTCTCATTTCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3959
 Db 541 CTGAGCCACAGGCTCTCATTTCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600
 Oy 3960 TTCCAAATCAGTTAGCTGTAGACATCTGTTGTAGAGAGTTGTTGCTCATTTTTCG 4019
 Db 601 TTCCAAATCAGTTAGCTGTAGACATCTGTTGTAGAGAGTTGTTGCTCATTTTTCG 660
 Oy 4020 TCTGTGCATTTTGTGAGACATTTGTGAGAAATTTCTATTTGGTCTCTACTGATTTT 4078
 Db 661 TCTGTGCATTTTGTGAGACATTTGTGAGAAATTTCTATTTGGTCTCTACTGATTTT 719

RESULT 5
 AK017582 1266 bp mRNA linear HTC 19-JAN-2002
 LOCUS AK017582
 DEFINITION Mus musculus 8 days embryo whole body CDNA, RIKEN full-length
 enriched library, clone:5730420B22:homolog to KIAA1344 PROTEIN
 (FRAGMENT), full insert sequence.

ACCESSION AK017582
 VERSION AK017582.1 GI:12856892
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (strain:C57BL/6J) 8 days embryo CDNA to mRNA,
 clone_lib:RIKEN full-length enriched mouse CDNA library
 clone:5730420B22.

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 Carninci,P. and Hayashizaki,Y.
 High-efficiency full-length CDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 99279253
 10349636

AUTHORS

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
 Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length CDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 20499374
 11042159

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
 Konno,H., Akiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M.,
 Sumi,N., Ishi,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
 Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
 Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
 Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,

TITLE
 Rikzen, Y., Muramatsu, M., Inoue, Y., Kita, A. and Haysashiak, Y.
 Open integrated sequence analysis (RISA) system--334-Format
 sequencing pipeline with 384 multiplexillary sequencer
 JOURNAL
 Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE
 20530913
 PUBMED
 11076861
 4
 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii,

| TITLE | AUTHORS |
|--|--|
| Functional annotation of a full-length mouse cDNA collection | Akawai, J., Shlnagawa, A., Shibata, R., Yoshino, M., Itoh, M., Ishii, Y. |
| Nature 409 (6821), 685-690 (2001) | Kawai, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, T., Fukuda, S. |
| | Atakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, T., Fukuda, S. |
| | Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I. |
| | Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., |
| | Kado, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., |
| | Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H., |
| | Kiehl, P., Lewis, S., Matsuo, Y., Nakado, I., Pesole, G., |
| | Konckenbush, J., Schriml, L. M., Stambli, F., Suzuki, R., Tomita, M., |
| | Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., |
| | Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., |
| | Carrinci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., |
| | Fletcher, C., Fujita, M., Gariboldi, M., Gustinchin, S., Hill, D., |
| | Hoffmann, M., Hume, D. A., Kamlay, M., Lee, N. H., Lyons, P., |
| | Marchionni, L., Mashima, J., Mazza, T., Mombaerts, P., Nordone, P., |
| | Ring, B., Ringwald, M., Rodriguez, S., Sakamoto, N., Sasaki, H., |
| | Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., |
| | Toyo-Oka, K., Wang, K. H., Weitz, C., Whitlaker, C., Wilming, L., |
| | Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohlsuk, S. |
| | and Hayashizaki, Y. |

| Db | 791 | TGCTCTTTTAGA | AAAAATATCGGCTCTGAGAGTATACAAATGCACATCTCTACTTTTTC | 850 |
|------------|--|---|---|--------------------|
| 0y | 965 | TTGTAACTAGTCTTTGGACCTTGACCCAGCATGTGAAAGACATAATGGAACAGCCATT | 1024 | |
| Db | 851 | TTGTAAACTGGCTTTGGACTTGACCCAGCATTTGCGAAGAACATAATGGAACAGCCATT | 910 | |
| 0y | 1025 | GACCTACAGTGAACATTCCTGTTTATTTAAAGCATATGAAGACACCTCTGTTGACTGAAGT | 1084 | |
| Db | 911 | GACCACTCATATATTCACGTAATTTGTGTAGACATATGAACGACCTCTGTTGATGAAAGT | 970 | |
| 0y | 1085 | TGCTGAAGATCCCTCAACAAGTTTCAAGTGTCCATCTCCAACTGGGCTTACCACACTGTTTT | 1144 | |
| Db | 971 | GCGTGAAGATCCCTCCAGAGGCTCTCAACCGTTCATCTACAACTAGATGGATGCCACTGTTTT | 1030 | |
| 0y | 1145 | TATTTGTAGCCAAAGCGCTACTTATGAAGCTATGAAAGAACTGCAAGATGGGTTGGCTTG | 1204 | |
| Db | 1031 | TATTTATGACCCAGCCGCGCTACCCAGAGGCTGTACAGAAAGAACTGGGAGATGGGTGCGTG | 1090 | |
| 0y | 1205 | GGCTTCTCTGGGAAAGAGAGAGTCTACTCTCTGTTTAAAGGACCTCTTTGGAAGTACAT | 1264 | |
| Db | 1091 | GCACCTCTCTGGGAAAGAGAGGCTCTCTCTCTGTTTGAAGGACCTCAATGATGTTAAACAT | 1150 | |
| 0y | 1285 | TCCCTCAAGATGTTATGTGTGCTTCCAAAAGACCAACAAGGACCTCCAGTGGAGATTTTTT | 1324 | |
| Db | 1151 | TCCCTGACAGCGTTATGTGTGCTTCCAAAAGACCAACAAGGATGTGCTGTGGAGATTTCT | 1210 | |
| 0y | 1325 | GGTATTCATCATGTTTGGATTTAAATATATTCATGTGGAATAATATATATGCACATTG | 1380 | |
| Db | 1211 | GGTGTGAATGATGTGTGATGTTATCATTTTCCACAGTGAATAATATATATGCACATTG | 1266 | |
| RESULT 6 | | | | |
| BM552780 | | 1453 bp | mRNA | linear |
| LOCUS | | | | EST 20-FEB-2002 |
| DEFINITION | BM552780 | AGENCOURT 6542342 NIH_MGC_119 | Homo sapiens | CDNA IMAGE:5742650 |
| ACCESSION | BM552780 | BM552780 | | |
| VERSION | BM552780.1 | GI:18790935 | | |
| KEYWORDS | EST. | | | |
| SOURCE | human. | | | |
| ORGANISM | Homo sapiens | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| AUTHORS | Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. | | | |
| TITLE | 1 (bases 1 to 1453) | | | |
| JOURNAL | NIH-MGC http://mgc.nci.nih.gov/ . | | | |
| COMMENT | National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs-remail.nih.gov Tissue Procurement: Life Technologies, Inc. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Plate: LLAM12761 row: C column: 03 High quality sequence stop: 628. Location/Qualifiers 1. 1453 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5742650" /clone_lid="NIH_MGC_119" /tissue_type="medulla" /lab_host="DH10B" /note="Organ: brain; Vector: pCMV-Sport6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for | | | |

| BASE COUNT | 458 a | 288 c | 284 g | 423 t | |
|-----------------------|--|---|------------|--------------|-----------------|
| ORIGIN | | | | | |
| Query Match | 14.1%; | Score 639.2; | DB 13; | Length 1453; | |
| Best Local Similarity | 99.4%; | Pred. No. 3.4e-11; | | | |
| Matches 652; | Conservative 0; | Mismatches 3; | Indels 1; | Gaps 1; | |
| QY | 2785 | CAAAAGTGACCTAATTTTATAGGGCTGTGTGTTCCAAAATTTTTTGGCATATAGACTT | 2844 | | |
| DB | 1 | CAAAAGTGACCTAATTTTATAGGGCTGTGTGTTCCAAAATTTTTTGGCATATAGACTT | 60 | | |
| QY | 2845 | AATTTATTTCCCTTAAAGAATATATTTAAATCATTTCAGTTGCGAGTACGACATCCA | 2904 | | |
| DB | 61 | AATTTATTTCCCTTAAAGAATATATTTAAATCATTTCAGTTGCGAGTACGACATCCA | 120 | | |
| QY | 2905 | ATAGATATTATTAATTAAGTACATATTTTATTTAAATTTCTAGTAAGTATTAACA | 2964 | | |
| DB | 121 | ATAGATATTATTAATTAAGTACATATTTTATTTAAATTTCTAGTAAGTATTAACA | 180 | | |
| QY | 2965 | AGCTAAAGTGGAGGGGCAAAATATTTTGGATTTTACTTTTCACCCAGTACATACCA | 3024 | | |
| DB | 181 | AGCTAAAGTGGAGGGGCAAAATATTTTGGATTTTACTTTTCACCCAGTACATACCA | 240 | | |
| QY | 3025 | AAATGCGAAATATAGAAATTTATTAATGAGATATTTTACATCCTTTTTTGACCAAGTCT | 3084 | | |
| DB | 241 | AAATGCGAAATATAGAAATTTATTAATGAGATATTTTACATCCTTTTTTGACCAAGTCT | 300 | | |
| QY | 3085 | TCTAATAGCAGACATATTTTACTTACTGATTTCTTACTTCGAGTAGGCATATTTTC | 3144 | | |
| DB | 301 | TCTAATAGCAGACATATTTTACTTACTGATTTCTTACTTCGAGTAGGCATATTTTC | 360 | | |
| QY | 3145 | AAGTGTTCATTTGCCACATGTGGCTGTGACTACTGTATTTGAGCAGTTCCAGTACAGACA | 3204 | | |
| DB | 361 | AAGTGTTCATTTGCCACATGTGGCTGTGACTACTGTATTTGAGCAGTTCCAGTACAGACA | 420 | | |
| QY | 3205 | AAACTAGCATTAATTAACCTTAGTTCAGGCATATTTCTATTGGATTTAAATTAACCTCT | 3264 | | |
| DB | 421 | AAACTAGCATTAATTAACCTTAGTTCAGGCATATTTCTATTGGATTTAAATTAACCTCT | 480 | | |
| QY | 3265 | AATTCAGTAACTGTCACAGTGCATTCATGACGTCAGCAGTATATTTGTTTATTTGGAG | 3324 | | |
| DB | 481 | AATTCAGTAACTGTCACAGTGCATTCATGACGTCAGCAGTATATTTGTTTATTTGGAG | 540 | | |
| QY | 3325 | TCATGATATTTAAATTCAGCGCTTTGTCAACCTCAGGGGATATTTAGCAATTTGCGGAGAC | 3384 | | |
| DB | 541 | TCATGATATTTAAATTCAGCGCTTTGTCAACCTCAGGGGATATTTAGCAATTTGCGGAGAC | 600 | | |
| QY | 3385 | ATTTTGATGTCATGACAGGCGACATTTATGACATTTAG-TCAGTAGAGGCGCATGG | 3439 | | |
| DB | 601 | ATTTTGATGTCATGACAGGCGACATTTATGACATTTAGTAGTAGAGGCGCAGGG | 656 | | |
| RESULT 7 | | | | | |
| LOCUS | AV646118 | 605 bp | mRNA | Linear | EST 15-JAN-2002 |
| DEFINITION | AV646118 | GLC Homo sapiens | CDNA clone | GLC/JR06 3', | mRNA sequence. |
| ACCESSION | AV646118 | | | | |
| VERSION | AV646118.1 | GI:9867132 | | | |
| KEYWORDS | EST. | | | | |
| SOURCE | human. | | | | |
| ORGANISM | Homo sapiens | | | | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| AUTHORS | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | | |
| TITLE | Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X., | | | | |
| | Xiao,H., Ou,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W., | | | | |
| | Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X., | | | | |
| | Hu,G., Gu,J., Chen,Z. and Han,Z. | | | | |
| | Insight into hepatocellular carcinogenesis at transcriptome level | | | | |
| | by comparing gene expression profiles of hepatocellular carcinoma | | | | |

JOURNAL
MEDLINE
COMMENT

with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
21625106
Contact: Zengqiang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzq@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
SOURCE

Location/Qualifiers
1. 605
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GLCAJF06"
/clone_lib="GLC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 173 a 131 c 156 g 142 t 3 others
ORIGIN

Query Match 13.3%; Score 602; DB 10; Length 605;
Best Local Similarity 99.7%; Pred. No. 4.5e-104;

Matches 602; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 GCCGAGTCGTGGCGGCTGACAGGCTGGAGGAGAGAGTCTACGCTTTGGCAGTTGGCG 60
1 GCCGAGTCGTGGCGGCTGACAGGCTGGAGGAGAGAGTCTACGCTTTGGCAGTTGGCG 60
61 AAGTGTTCACAGGCTACCCGGCTAGTCTGGCAGCGGCCCGCTTTCCTCTCTCTCT 120
61 AAGTGTTCACAGGCTACCCGGCTAGTCTGGCAGCGGCCCGCTTTCCTCTCTCTCT 120
121 CGCGTGGCGGGGAGAACTGTTGGCGCGCGGCTCGGAGACGGCCAGCTCCCGCGCGC 180
121 CGCGTGGCGGGGAGAACTGTTGGCGCGCGGCTCGGAGACGGCCAGCTCCCGCGCGC 180
181 AGGTCCCGGCGAGTAACATAGATCATCAGTAGAGAACTTCTGTAAGTTGTTCAAGAAA 240
181 AGGTCCCGGCGAGTAACATAGATCATCAGTAGAGAACTTCTGTAAGTTGTTCAAGAAA 240
241 ATTGGAAGTAGCAAAATAGAAATTAAGAAATTAACACAGATACAGAGACAGCATGGA 300
241 ATTGGAAGTAGCAAAATAGAAATTAAGAAATTAACACAGATACAGAGACAGCATGGA 300
301 AGTGTGCTTAGGAAACAGAACAGACAGAGTGAAGAAATTCGCTCGATGAC 360
301 AGTGTGCTTAGGAAACAGAACAGACAGAGTGAAGAAATTCGCTCGATGAC 360
361 AACTGACAGCTATATGTTTCCGGCTTCAATGCTTTAGATTGGATCTCTTTTCA 420
361 AACTGACAGCTATATGTTTCCGGCTTCAATGCTTTAGATTGGATCTCTTTTCA 420
421 AACTGACAGCTATATGTTTCCGGCTTCAATGCTTTAGATTGGATCTCTTTTCA 420
421 AACTGACAGCTATATGTTTCCGGCTTCAATGCTTTAGATTGGATCTCTTTTCA 420
421 TAAATGATTTTACATGCGACAGTAAACTCTTTACCGAGACTGAGTCTCAGAAAT 480
421 TAAATGATTTTACATGCGACAGTAAACTCTTTACCGAGACTGAGTCTCAGAAAT 480
481 ATTTAGTACATTTGCAACACAGCTCTTGAAGAACTGAAGAACTGAAGAACTGAAG 540
481 ATTTAGTACATTTGCAACACAGCTCTTGAAGAACTGAAGAACTGAAGAACTGAAG 540
481 ATTTAGTACATTTGCAACACAGCTCTTGAAGAACTGAAGAACTGAAGAACTGAAG 540
541 ACTATGAAATTTGCAAGGTTAAATTTGTTCAAGAAATATATCAAGATCTGTG 600
541 ACTATGAAATTTGCAAGGTTAAATTTGTTCAAGAAATATATCAAGATCTGTG 600
601 GAAA 604
601 GAAA 604

RESULT 8
B1712516/c
LOCUS
DEFINITION
B1712516 593 bp mRNA linear EST 11-MAR-2002
1e08h05.x1 HR85 islet Homo sapiens CDNA clone IMAGE:5086689 3',
mRNA sequence.

ACCESSION
B1712516
VERSION
B1712516.1 GI:15688211
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS

1 (bases 1 to 593)
Mellon, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Keastner, K.,
Lemishka, L., Scaer, M., Brestelli, J., Gradow, G., Clifton, S.,
Hillier, L., Ma, M., Page, D., Wyle, T., Martin, J., Bistabin, A.,
Schmitt, A., Theising, B., Rittler, E., Ronko, J., Bennett, J., Cardenas,
M., Gibbons, M., McCann, R., Cole, R., Tsagarelis, R., Williams, T.,
Jackson, Y., and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Mellon, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmellon@hlohp.harvard.edu

Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
Washington University Genome Sequencing Center for information on
obtaining a clone please contact: Dr. Hiroshi Inoue
(hinoue@im.wustl.edu)
Seq primer: -400p from Gibco
High quality sequence stop: 430.

FEATURES
SOURCE

Location/Qualifiers
1. 593
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5086689"
/clone_lib="HR85 islet"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
NotI; Site_2: XhoI; CDNA made by oligo-dT priming.
Size-selected on agarose gel. Average insert size ~1kb. 5'
XhoI site was destroyed after directional cloning.
Amplified once. Contact information: Hiroshi Inoue, MD,
Metabolism Div. (Alan Permutt lab), Washington University
School of Medicine, Box 8127, 660 South Euclid Ave., St.
Louis, MO 63110, E-mail: hinoue@im.wustl.edu, Tel:
314-362-1916, Fax: 314-747-2692."

BASE COUNT 237 a 93 c 91 g 172 t
ORIGIN

Query Match 12.9%; Score 583.6; DB 13; Length 593;
Best Local Similarity 99.0%; Pred. No. 1.4e-100;
Matches 586; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

3916 CATTCGTGAGGAGGAGAAAGTATGCAAGAGTAAATTTTCATCCAAATCACTAGC 3975
3916 CATTCGTGAGGAGGAGAAAGTATGCAAGAGTAAATTTTCATCCAAATCACTAGC 3975
593 CATTCGTGAGGAGGAGAAAGTATGCAAGAGTAAATTTTCATCCAAATCACTAGC 534
593 CATTCGTGAGGAGGAGAAAGTATGCAAGAGTAAATTTTCATCCAAATCACTAGC 534
3976 TGTAGACTGATCTGTTGACAGCTTTGCTCATTTTGTCTGCTGATTTTGA 4035
3976 TGTAGACTGATCTGTTGACAGCTTTGCTCATTTTGTCTGCTGATTTTGA 4035
533 TGTAGACTGATCTGTTGACAGCTTTGCTCATTTTGTCTGCTGATTTTGA 474
533 TGTAGACTGATCTGTTGACAGCTTTGCTCATTTTGTCTGCTGATTTTGA 474
4036 GACATTTTGTGAGATATTTGCTGCTCTACTGATATTTTCTTTTAATATCTACT 4095
4036 GACATTTTGTGAGATATTTGCTGCTCTACTGATATTTTCTTTTAATATCTACT 4095
473 GACATTTTGTGAGATATTTGCTGCTCTACTGATATTTTCTTTTAATATCTACT 414
473 GACATTTTGTGAGATATTTGCTGCTCTACTGATATTTTCTTTTAATATCTACT 414
4096 TGATATCTTGTCTTAATTTTCTTACATATGTTTGCCTGATACAGATTTTAT 4155
4096 TGATATCTTGTCTTAATTTTCTTACATATGTTTGCCTGATACAGATTTTAT 4155

```

|||||
Db 413 TGAATACCTGTTCTTAAATTTCTTACATATGGTTTGGCTGATCAACTGATTTTAT 354
OY 4156 AACTGAAATTTAAGATCATCAGCTAAACAGTAAAGTGCATATTTCCCTTTAAC 4215
Db 353 AACTGAAATTTAAGATTTAAGACGTAAACAGTAAAGTGCATATTTCCCTTTAAC 294
OY 4216 ATAGACCCGTTGCTACTCTCAGACCCCTCTCTCAATTTTTTCTGTAGCATGTGATG 4275
Db 293 ATAGACCCGTTGCTACTCTCAGACCCCTCTCTCAATTTTTTCTGTAGCATGTGATG 234
OY 4276 CCGATTAATACATTTTATTTGCTTTATTTCTAATATAGGAACAATAGAGTGAAT 4335
Db 233 CCGATTAATACATTTTATTTGCTTTATTTCTAATATAGGAACAATAGAGTGAAT 174
OY 4336 CTAATATATAGTTGTATATATAACATCATAGCTTAATTTATAGAAATGCTAATTA 4395
Db 173 CTAATATATAGTTGTATATATAACATCATAGCTTAATTTATAGAAATGCTAATTA 114
OY 4396 GTACACGACACATAGAAACATGAATTTGCTTATGTCATTTGTCAGCAATTTTGAC 4455
Db 113 GTGCGACGACATAGAAACATGAATTTGCTTATGTCATTTGTCAGCAATTTTGAC 54
OY 4456 AGTCATTAATGTTTGTATATATTTTAAATTAAGTGTCTGGCTTCCAGATAC 4507
Db 53 AGTCATTAATGTTTGTATATATTTTAAATTAAGTGTCTGGCTTCCAGATAC 2

RESULT 9
AM961708 612 bp mRNA linear EST 01-JUN-2000
LOCUS EST373781 MAGE resequences, MAGC Homo sapiens cDNA, mRNA sequence.
ACCESSION AM961708
VERSION AM961708.1 GI:8151392
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 612)
Hegde,P., Qi,R., Abernathy,K., Dharp,S., Gaspard,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
Unpublished (2000)
JOURNAL The Institute for Genomic Research
COMMENT 9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@ligr.org
Plate: 164
Seq primer: Reverse.
FEATURES
Source Location/Qualifiers
1..612
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGC"
/note="Vector: pBluescriptSkm"
BASE COUNT 202 a 119 c 116 g 173 t 2 others
ORIGIN
Query Match 12.8%; Score 581.2; DB 10; Length 612;
Best Local Similarity 98.9%; Pred. No. 4e-100;
Matches 605; Conservative 0; Mismatches 5; Indels 2; Gaps 2;
OY 1912 GACTATTTAGTCCACCATGAAACAGCAAAAGAGATTTTATGTAGAGCAGAACTACC 1971
Db 1 GACTATTTAGTCCACCATGAAACAGCAAAAGAGATTTTATGTAGAGCAGAACTACC 60
OY 1972 TAAAGGATATGTTATCACTGGAATTTATTTGAGACAGATGTTTGTACTGTCAACA 2031
|||||

```

```

Db 61 TAAAGGATATGTTATCACTGGAATTTATTTGAGAGAGATGTTTGTACTGTCAACA 120
OY 2032 AATATGCTGCAAGCTTTCCAGCCCTGCTGCTCCAGACACAGAGCAAAATAGAGA 2091
Db 121 AATATGCTGCAAGCTTTCCAGCCCTGCTGCTCCAGACACAGAGCAAAATAGAGA 180
OY 2092 GCATCCCACTAGCTAGACACATGACACAGACATAGTTCAAATATTAACAGATGACTAC 2151
Db 181 GCATCCCACTAGCTAGACACATGACACAGACATAGTTCAAATATTAACAGATGACTAC 240
OY 2152 TGGAAATGTTCCGGAATCACTGTGAAATGCTCCAGTTTTCAGACTCAGAAAC 2211
Db 241 TGGAAATGTTCCGGAATCACTGTGAAATGCTCCAGTTTTCAGACTCAGAAAC 300
OY 2212 CATTATTTGATTTTGTTCAGTATGACATGTAAATCCTCAATATTAAGCAATTTGA 2271
Db 301 CATTATTTGATTTTGTTCAGTATGACATGTAAATCCTCAATATTAAGCAATTTGA 360
OY 2272 CACTGTTAAAGCAGAAATACCTTGATTCATTTACTCAGCTGTGTTAAATCTTAAGAAAT 2331
Db 361 CACTGTTAAAGCAGAAATACCTTGATTCATTTACTCAGCTGTGTTAAATCTTAAGAAAT 420
OY 2332 CTCACGTGGGAGAGGAATCTGCGGCATATTTTGATCCTGCGCTCCCTTCCTTC 2391
Db 421 CTCACGTGGGAGAGGAATCTGCGGCATATTTTGATCCTGCGCTCCCTTCCTTC 480
OY 2392 TTGTTTGTGTAATCTGCATTCAGTGTGCGCAAGTATTTTCCCTTCCAGACAGGCTA 2451
Db 481 TTGTTTGTGTAATCTGCATTCAGTGTGCGCAAGTATTTTCCCTTCCAGACAGGCTA 540
OY 2452 TAATTTAGAGAAACCTTTGATTTGCTGTGAG- AAATTGAGACAGAGAC-TAGAAATCA 2509
Db 541 TAATTTAGAGAAACCTTTGATTTGCTGTGAG- AAATTGAGACAGAGACTTAGAAATCA 600
OY 2510 TATCACAATTTT 2521
Db 601 TATCACAATTTT 612

RESULT 10
BG928131 686 bp mRNA linear EST 06-NOV-2001
LOCUS HNC65-1-B8.R.R HNC (Human Normal Cartilage) Homo sapiens cDNA, mRNA
SEQUENCE.
ACCESSION BG928131
VERSION BG928131.1 GI:14322654
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 686)
Kumar,S., Connor,J.R., Dadds,R.A., Halsey,W., Van Horn,M., Mao,J.,
Sathre,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
Lark,M.W.
Identification and initial characterization of 5000 expressed
sequenced tags (ESTs) each from adult human normal and
osteoarthr. Cartil. 9 (7), 641-653 (2001)
JOURNAL 21482651
MEDLINE Contact: Sanjay Kumar
COMMENT UW2109
GlaxoSmithKline
709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
Tel: 610-270-7245
Fax: 610-270-5598
Email: sanjay_kumar-1@sk.com
Seq primer: 17.
FEATURES
Source Location/Qualifiers
1..686
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HNC (Human Normal Cartilage)"

```

```

/tissue_type="cartilage"
/lab_host="E.coli DH10 B"
/Note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
Directional"
BASE COUNT      188 a      132 c      127 g      239 t
ORIGIN
Query Match      12.6%: Score 568.2; DB 13; Length 686;
Best Local Similarity 99.9%; Pred. No. 1.1e-97;
Matches 581; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 3589 CATGTTGAAAAAATTTGGCTATTAAAGATGATTTAGATGGCTTATCTGATTTAT
DB 86 CATGTTGAAAAAATTTGGCTATTAAAGATGATTTAGATGGCTTATCTGATTTAT 145
QY 3649 ACCGAGTACACCTGATCTTTCTAATTTTTCAGAAAGTGAAGGATTAACCTAGAG
DB 146 ACCGAGTACACCTGATCTTTCTAATTTTTCAGAAAGTGAAGGATTAACCTAGAG 205
QY 3709 AGAGCTAGAGATGATTTATTTTAAAGTGAAGCTTAAACCTCTTATTTCTACAA
DB 206 AGAGCTAGAGATGATTTATTTTAAAGTGAAGCTTAAACCTCTTATTTCTACAA 265
QY 3769 GTTATATGGCTAAATTTTCAGATGACAGGATTCAGCATTCCTCATGAGAA 3828
DB 266 GTTATATGGCTAAATTTTCAGATGACAGGATTCAGCATTCCTCATGAGAA 325
QY 3829 AGAAGAGCTCCCTCATCTGAAAGCTCTGAAATCTACCTTGCAGCTTCAGACAATC
DB 326 AGAAGAGCTCCCTCATCTGAAAGCTCTGAAATCTACCTTGCAGCTTCAGACAATC 385
QY 3889 AGTGTATCTCCCTGAGCAGCAGGCTCATCTGTGAGGAGGAGAAAGATTAGCCAAAGA
DB 386 AGTGTATCTCCCTGAGCAGCAGGCTCATCTGTGAGGAGGAGAAAGATTAGCCAAAGA 445
QY 3949 GTTATATTTTCATCCAAATTCATAGCTGTAGACTGATCTGTTTGTAGACAGTTGTTGT
DB 446 GTTATATTTTCATCCAAATTCATAGCTGTAGACTGATCTGTTTGTAGACAGTTGTTGT 505
QY 4009 CTCATTTTGTCTGCTGATTTTGTGAGACATTTGTGAGAAATATCTTATTTGGTCTCT
DB 506 CTCATTTTGTCTGCTGATTTTGTGAGACATTTGTGAGAAATATCTTATTTGGTCTCT 565
QY 4069 ACTGTATTTTCTTTTAAATATCTACTGATGATCTGTTCTTAAATTTTCTACATAT
DB 566 ACTGTATTTTCTTTTAAATATCTACTGATGATCTGTTCTTAAATTTTCTTACATAT 625
QY 4129 GG-TTGGCTGATACAACTGATTTTATTACTGAAATTTTAAGAA 4172
DB 626 GGTTTGGCTGATACAACTGATTTTATTACTGAAATTTTAAGGA 670

RESULT 11
BM972493/c      585 bp      mRNA      linear      EST 21-MAR-2002
LOCUS           UI-CF-EC1-abq-p-23-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone
DEFINITION      UI-CF-EC1-abq-p-23-0-UI 3', mRNA sequence.
ACCESSION       BM972493
VERSION         BM972493.1 GI:19590079
KEYWORDS        EST.
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE       1 (bases 1 to 585)
AUTHORS         Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE           Normalization and subtraction: two approaches to facilitate gene
JOURNAL         Genome Res. 6 (9), 791-806 (1996)
MEDLINE         97044477
COMMENT         Contact: McCray, PB
                  McCray Lab
                  University of Iowa

```

```

2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul.mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source
location/Qualifiers
1..585
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-CF-EC1-abq-p-23-0-UI"
/clone_1lb="UI-CF-EC1"
/tissue_type="Lung"
/dev_stage="Adult and Fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/Note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-CF-EC1 is a normalized cDNA library containing the
following tissue(s): Normal lung from adult and from fetal
day 64, day 87, week 19 and week 42. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
TAG-LIB=UI-CF-EC1
TAG-TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
and 380-383
TAG_SEQ=AAAGCTGTAC"
BASE COUNT      228 a      83 c      92 g      182 t
ORIGIN
Query Match      12.5%: Score 567; DB 14; Length 585;
Best Local Similarity 97.9%; Pred. No. 2e-97;
Matches 573; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 3933 AAAGATTAGCCAAAGAGTTAATTTTCATTCCAATCACTAGCTGATGATCTGT 3992
DB 585 AAAGATTAGCCAAAGAGTTAATTTTCATTCCAATCACTAGCTGATGATGATCTGT 526
QY 3993 TGTAGCACTGTTGTCTCATTTTGTGCTGTGCAATTTTGTGAGCAATTTGTTGAATA 4052
DB 525 TGTAGCACTGTTGTCTCATTTTGTGCTGTGCAATTTTGTGAGCAATTTGTTGAATA 466
QY 4053 TTCTATTTGGTCTACTGATTTTCTTTTAAATATCTACTGATGATCTGTTCTTTA 4112
DB 465 TTCTATTTGGTCTACTGATTTTCTTTTAAATATCTACTGATGATCTGTTCTTTA 406
QY 4113 AATTTTCTTCACATATGTTGGCTGATACAACTGATTTTATTAACGAAATTTAAGAA 4172
DB 405 AATTTTCTTCACATATGTTGGCTGATACAACTGATTTTATTAACGAAATTTAAGAA 346
QY 4173 TCTAAGACCTAAACTGAGTAGTGCATATTTCTTATAACATAGACCGTGTCTACT 4232
DB 345 TCTAAGACCTAAACTGAGTAGTGCATATTTCTTATAACATAGACCGTGTCTACT 286
QY 4233 CTCAGACACCTCTCTCATTTTCTTCTGATGATGATGATGATGATGATGATGATGAT 4292
DB 285 CTCAGACACCTCTCTCATTTTCTTCTGATGATGATGATGATGATGATGATGATGAT 226
QY 4293 TCATTTGCTTTTATTTTAAATAGGAACTAGAGAGTGAACCTAAATATAGTTGATG 4352

```

```
Db 225 TCATTGCTTTTATTTCTAATATGGACACATGAGTGACTCTAATATAGCTGTG 166
QY 4353 TATATAACATCATAGTACCTAATTTATTAGAAATGCTAATTAAGTACGACATAGAAA 4412
Db 165 TATATAACATCATAGTACCTAATTTATTAGAAATGCTAATTAAGTACGACATAGAAA 106
QY 4413 CATGAATATGCTTACATTTGACCTTTTGTGACCAATTTTGACAGTCATTAATGTTTGC 4472
Db 105 CATGAATATGCTTACATTTGACCTTTTGTGACCAATTTTGACAGTCATTAATGTTTGC 46
QY 4473 ATATATTTTAAATTAAGTGTGCTGGTTTCAGAAATACCTCAAAAAA 4517
Db 45 ATATATTTTAAATTAAGTGTGCTGGTTTCAGAAAAA 1

RESULT 12
AL040840/c 564 bp mRNA linear EST 29-FEB-2000
LOCUS DKEP434F0915_s1 434 (synonym: htes3) Homo sapiens cDNA clone
DEFINITION DKEP434F0915_3', mRNA sequence.
ACCESSION AL040840
VERSION AL040840
KEYWORDS GI:5409785
SOURCE EST.
ORGANISM human.
REFERENCE Homo sapiens
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 564)
Bloeker, H., Boecker, M., Brandt, P., Mewes, H.W., Gassenhuber, J. and
Wiemann, S.
EST (Bioecker, et al.)
JOURNAL Unpublished (1999)
COMMENT Contact: Bloecker H
MIPS
Am Klopfersplitz 18a D-82152 Martinsried, Germany
This is the 3' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No r1 sequence available.
This clone (DKEP434F0915) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
location/Qualifiers
source 1..564
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKEP434F0915"
/clone_lib="434 (synonym: htes3)"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI"

BASE COUNT 221 a 88 c 95 g 160 t
ORIGIN
Query Match 12.5%; Score 563.6; DB 9: Length 564;
Best Local Similarity 99.8%; Pred. No. 8.8e-97;
Matches 563: Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3910 CGGCTCATCTGTGAGGAGGAGAAAGATTAGCAAGATTAAATTTTCATTCACAAATCA 3969
Db 564 CGGCTCATCTGTGAGGAGGAGAAAGATTAGCAAGATTAAATTTTCATTCACAAATCA 505
QY 3970 CTTAGCTGTAGACTGATCTGTTGTAGCAGTTGTTGCTCATTTTTCCTGTCGTCATT 4029
Db 504 CTTAGCTGTAGACTGATCTGTTGTAGCAGTTGTTGCTCATTTTTCCTGTCGTCATT 445
QY 4030 TTTTGAAGATTTTGAAGAATATCTATTTGGTCTCTACACTATTTTCTTTTAATA 4089
Db 444 TTTTGAAGATTTTGAAGAATATCTATTTGGTCTCTACACTATTTTCTTTTAATA 385
```

```
QY 4090 TCTACTGTATATCTGTCTTTAAATTTTCTTCACATATGTTTGGCTGATACACTGAT 4149
Db 384 TCTACTGTATATCTGTCTTTAAATTTTCTTCACATATGTTTGGCTGATACACTGAT 325
QY 4150 TTTTATACTGAATTTAAGAAATCTACACAGTAAACACAGTAAGTATATTTTCCCT 4209
Db 324 TTTTATACTGAATTTAAGAAATCTACACAGTAAACACAGTAAGTATATTTTCCCT 265
QY 4210 TATACATAGACCCGTTGCTACTCTCAGACCCCTCTCAATTTTTCCTGTAGCAT 4269
Db 264 TATACATAGACCCGTTGCTACTCTCAGACCCCTCTCAATTTTTCCTGTAGCAT 205
QY 4270 GTGATGCCGTGATTTAAACATATTTTCATTTGCTTTTAAATTTGGAACATGAGAG 4329
Db 204 GTGATGCCGTGATTTAAACATATTTTCATTTGCTTTTAAATTTGGAACATGAGAG 145
QY 4330 TGACCTCTAATTTAGTGTGCTAGTAAACATCTTTGCGCTAATTTATAGAAATGCT 4389
Db 144 TGACCTCTAATTTAGTGTGCTAGTAAACATCTTTGCGCTAATTTATAGAAATGCT 85
QY 4390 AATTAGTACACAGCATAGAAACATGAATGCTTAGTCAATTTGCTTGTACAGCAAT 4449
Db 84 AATTAGTACACAGCATAGAAACATGAATGCTTAGTCAATTTGCTTGTACAGCAAT 25
QY 4450 TTTGACAGTCATTAATGTTTGTCA 4473
Db 24 TTTGACAGTCATTAATGTTTGTCA 1

RESULT 13
BE895218 900 bp mRNA linear EST 20-OCT-2000
LOCUS 60143654F1 NIH_MGC_72 Homo sapiens cDNA IMAGE:391887 5',
DEFINITION mRNA sequence.
ACCESSION BE895218
VERSION BE895218.1 GI:10358391
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 900)
NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC/DCFD/DTF
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM9747 row: 1 column: 14
High quality sequence scop: 557.
FEATURES
location/Qualifiers
source 1..900
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3918877"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT 252 a 155 c 177 g 315 t 1 others
ORIGIN
Query Match 12.4%; Score 562; DB 12: Length 900;
```


Best Local Similarity 90.6%; Pred. No. 1.6e-96;
Matches 645; Conservative 1; Mismatches 56; Indels 10; Gaps 4;

```

Qy 3652 TGGATACAACTGATCTTTCTAATATTTTCAGAAAGATGATGGGATTAACCCCTAGAGAG 3711
Db 1 TGGATACAACTGATCTTTCTAATATTTTCAGAAAGATGATGGGATTAACCCCTAGAGAG 60
Qy 3712 ACTCAGATGATATTTATTTTAAAGTGAAGTCTTAAACCTCTCTTATTTTACAGAT 3771
Db 61 ACTCAGATGATATTTATTTTAAAGTGAAGTCTTAAACCTCTCTTATTTTACAGAT 120
Qy 3772 ATATGGCTAAATTTTCAGATTTGAACAGGATTCAGATTCCTGCAATCTCTCATGAGAGA 3831
Db 121 ATATGGCTAAATTTTCAGATTTGAACAGGATTCAGATTCCTGCAATCTCTCATGAGAGA 180
Qy 3832 GAGGCTCCCTCATCTGAGAGCGCTCTGAAATCTACCCCTTGCAAGCTTCAGAAATCACT 3891
Db 181 GAGGCTCCCTCATCTGAGAGCGCTCTGAAATCTACCCCTTGCAAGCTTCAGAAATCACT 240
Qy 3892 TGATCTCCCTGAGCCACAGCGCTCATCTGAGAGAGGAGAAAGATTAGCCAAAGAGT 3951
Db 241 TGATCTCCCTGAGCCACAGCGCTCATCTGAGAGAGGAGAAAGATTAGCCAAAGAGT 300
Qy 3952 AATTTTCATCCAAATCACTTACCTGATGATGATGATGATGATGATGATGATGATGAT 4011
Db 301 AATTTTCATCCAAATCACTTACCTGATGATGATGATGATGATGATGATGATGATGAT 360
Qy 4012 ATTTTGTCTGCTGATTTTTCAGACATTTTTCAGACATTTTTCAGACATTTTTCAGACAT 4071
Db 361 ATTTTGTCTGCTGATTTTTCAGACATTTTTCAGACATTTTTCAGACATTTTTCAGACAT 420
Qy 4072 GATTTTCTTTTAAATCTACTGATGATGATGATGATGATGATGATGATGATGATGAT 4131
Db 421 GATTTTCTTTTAAATCTACTGATGATGATGATGATGATGATGATGATGATGATGAT 480
Qy 4132 TTGCTGATACAACTGATTTTTCAGAAATTTTTCAGAAATTTTTCAGAAATTTTTCAG 4191
Db 481 TTGCTGATACAACTGATTTTTCAGAAATTTTTCAGAAATTTTTCAGAAATTTTTCAG 539
Qy 4192 TAAAGGCAATATTTTCTTATTAATGATGATGATGATGATGATGATGATGATGATGAT 4251
Db 540 TAAAGGCAATATTTTCTTATTAATGATGATGATGATGATGATGATGATGATGATGAT 595
Qy 4252 TTTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4311
Db 596 TTTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 650
Qy 4312 ATATGGCAAACTGAGAGTGAATCTTAAATATAGTTGATGATTAATTAACAT 4363
Db 651 TCTGTTGGAGCCCTGGAGGAGCCCTTAATTTGGGTGGTATTAATAAACATCTT 702

```

RESULT 14
BF790996 695 bp mRNA linear EST 12-JAN-2001
LOCUS 602251048F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4338659 5',
DEFINITION mRNA sequence.
ACCESSION BF790996
VERSION BF790996.1 GI:12096050
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 695)
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps@email.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1C41214 row: 0 column: 12
High quality sequence stop: 581.
Location/Qualifiers

```

FEATURES
source
1..695
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4338659"
/clone_lib="NIH_MGC_81"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: muscle (skeletal); Vector: pDNR-LTR
(Clontech); Site: 1: SfiI (ggcgccggcgcc); Site: 2: SfiI
(ggcatatggcc); 5' and 3' adaptors were used in cloning
as follows: 5' adaptor sequence:
5'-ATTCTAGAGGCGGAGCGGCGGAGCATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size
1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
inserts by PCR. This library was enriched for full-length
clones and was constructed by Clontech Laboratories (Palo
Alto, CA)."
BASE COUNT 199 a 134 c 130 g 232 t
ORIGIN

```

Query Match 12.3%; Score 556.4; DB 12; Length 695;
Best Local Similarity 94.5%; Pred. No. 2e-95;
Matches 641; Conservative 0; Mismatches 31; Indels 6; Gaps 6;

```

Qy 3522 TAACCTTGTTTAAAGTAACTGTCGAGACATGTTCCAGAAATTTCTGCTGCT 3581
Db 1 TAACCTTGTTTAAAGTAACTGTCGAGACATGTTCCAGAAATTTCTGCTGCT 60
Qy 3582 CACGATATCATGTTTGAAGAAATTTGGCTATTAAAGATATGATGATGATGAT 3641
Db 61 CACGATATCATG-TTGAAGAAATTTGGCTATTAAAGATATGATGATGATGAT 119
Qy 3642 GATTTATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3701
Db 120 GATTTATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 179
Qy 3702 CTAG-AAGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3760
Db 180 CTAGCAAGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 239
Qy 3761 TTCTACAACTTATGAGCTAAATTTTCAGATGATGATGATGATGATGATGATGAT 3820
Db 240 TTCTACAACTTATGAGCTAAATTTTCAGATGATGATGATGATGATGATGATGAT 299
Qy 3821 TCATGGAAGAGAGGCTCCCTCATCTGAGAGCGCTCTGAAATCTACCTTGAAGCTTCA 3880
Db 300 TCATGGAAGAGAGGCTCCCTCATCTGAGAGCGCTCTGAAATCTACCTTGAAGCTTCA 359
Qy 3881 GACAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3940
Db 360 GACAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 419
Qy 3941 GCCAAAGAGTAAATTTTCATTCCTAGCTGTTAGACTGATGATGATGATGATGAT 4000
Db 420 GCCAAAGAGTAAATTTTCATTCCTAGCTGTTAGACTGATGATGATGATGATGAT 478
Qy 4001 TTGTTGCTCATTTTTCCTGCTGCTGATGATGATGATGATGATGATGATGATGAT 4060
Db 479 TTGTTGCTCATTTTTCCTGCTGCTGATGATGATGATGATGATGATGATGATGAT 537
Qy 4061 GGTGCTACCTGATTTTCTTTAATATCTGATGATGATGATGATGATGATGATGAT 4120
Db 538 GGTGCTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 597
Qy 4121 TCACATATGTTGGCTGATCACTGATTTTATTAAGTAATTAAGTAATTAAGTA 4180
Db 111 TCACATATGTTGGCTGATCACTGATTTTATTAAGTAATTAAGTAATTAAGTA 111

```

Db 598 TCAATTATGGCTTGCCCTGATA-AACGTGATTCTATACTGGAATCTCAGCAAT-TAACAG 655
 QY 4181 CTAAACTCAGTAAGTGC 4198
 Db 656 CTAAACTCAGTAGGGGC 673

RESULT 15 AM65394/c

LOCUS hi90c01.x1 Soares_NFL.T.GBC.S1 Homo sapiens mRNA linear EST 06-APR-2000
 DEFINITION IMAGE:2979552 3', mRNA sequence.

ACCESSION AM65394
 VERSION AM65394.1 GI:7457940

KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LNLN; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.

Seq primer: -40UP from Gibco
 High quality sequence stop: 462.

FEATURES
 Location/Qualifiers

1..556
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2979552"
 /clone_lib="Soares_NFL.T.GBC.S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT7T3P-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NBHL19W, testis NHT, and B-cell
 NCI CGAP GC81) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 I.M.G.E. clones 297480-302087, 682632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT 224 a 80 c 90 g 161 t 1 others
 ORIGIN

Query Match 12.2%; Score 552.6; DB 10; Length 556;
 Best Local Similarity 99.6%; Pred. No. 1.1e-94;
 Matches 552; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3953 ATTTTCATTCACAAATCAGTCTAGCTGTAGACGATCTGTTGTAGCAGTGTGTTGTCTCA 4012
 Db 556 ATTTTCATTCACAAATCAGTCTAGCTGTAGACGATCTGTTGTAGCAGTGTGTTGTCTCA 497
 QY 4013 TTTTTCCTGTCGTCATTTTGGAGACATTTGGAGAAATTTCTATTGGTGTCTACTG 4072
 Db 496 TTTTTCCTGTCGTCATTTTGGAGACATTTGGAGAAATTTCTATTGGTGTCTACTG 437
 QY 4073 TATTTTCTTTTAATATCTACTGATATCTGTCTTTAAATTTTCTTCACATATGTT 4132
 Db 436 TATTTTCTTTTAATATCTACTGATATCTGTCTTTAAATTTTCTTCACATATGTT 377
 QY 4133 TGGCTGATACACTGATTTTATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 4192
 Db 376 TGGCTGATACACTGATTTTATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 317
 QY 4193 AAGTGATMTATTTCTTATACATAGACCCGTTGCTACTCTCAGACCCCTCTCTCAAT 4252

Db 316 AAGTGATMTATTTCTTATACATAGACCCGTTGCTACTCTCAGACCCCTCTCTCAAT 257
 QY 4253 TTTTTCCTGTCGTCATGTGATGCTGATTAACCTCATTTTTCATTTGCTTTTATTTCTAA 4312
 Db 256 TTTTTCCTGTCGTCATGTGATGCTGATTAACCTCATTTTTCATTTGCTTTTATTTCTAA 197
 QY 4313 TATGGACAAATGAGAGTGAACCTCTAAATATAGTTGTAGTAAATTAACATCATTAAGCCT 4372
 Db 196 TATGGACAAATGAGAGTGAACCTCTAAATATAGTTGTAGTAAATTAACATCATTAAGCCT 137
 QY 4373 AATTATTAAGAAATGCTAAATTAAGTACCAACATGAATGCTTAAGTAT 4432
 Db 136 AATTATTAAGAAATGCTAAATTAAGTACCAACATGAATGCTTAAGTAT 77
 QY 4433 GTACCTTTGTCAGCAATTTTGAACAGTCATTAATGTTTGCATTAATTAATTAAGTAT 4492
 Db 76 GTACCTTTGTCAGCAATTTTGAACAGTCATTAATGTTTGCATTAATTAATTAAGTAT 17
 QY 4493 TGGGTTTCAGAAATA 4506
 Db 16 TGGGTTTCAGAAATA 3

Search completed: March 22, 2003, 10:47:45
 Job time: 6639 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:10:12 ; Search time 39 Seconds
(without alignments)
2757.263 Million cell updates/sec

Title: US-09-847-046-2
Perfect score: 4142
Sequence: 1 MFSGFNFRVGISFVIMCIF.....FKFAEKSPRDEKELGCSKVN 807

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*

1: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/genescp-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/genescp-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/genescp-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/genescp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 4142 | 100.0 | 807 | 23 | ABG61812 |
| 2 | 1839 | 44.4 | 357 | 22 | AAB94073 |
| 3 | 185 | 4.5 | 485 | 15 | AAR49829 |
| 4 | 185 | 4.5 | 505 | 16 | AAR73963 |
| 5 | 180 | 4.3 | 505 | 15 | AAR63807 |
| 6 | 177 | 4.3 | 496 | 22 | ABB62855 |
| 7 | 175.5 | 4.2 | 505 | 16 | AAR82801 |
| 8 | 172.5 | 4.2 | 511 | 16 | AAR65506 |
| 9 | 172.5 | 4.2 | 515 | 16 | AAR65504 |
| 10 | 172.5 | 4.2 | 515 | 22 | AAB49946 |

| | | | | | |
|----|-------|-----|------|----|----------|
| 11 | 169.5 | 4.1 | 515 | 22 | AAE12991 |
| 12 | 169.5 | 4.1 | 515 | 22 | AAB82539 |
| 13 | 169.5 | 4.1 | 544 | 21 | AAB43424 |
| 14 | 169.5 | 4.1 | 549 | 23 | ABP41717 |
| 15 | 166.5 | 4.0 | 505 | 16 | AAR70329 |
| 16 | 166.5 | 4.0 | 515 | 16 | AAR69505 |
| 17 | 164 | 4.0 | 454 | 22 | AAB49346 |
| 18 | 155 | 3.7 | 489 | 21 | AAB49944 |
| 19 | 153.5 | 3.7 | 489 | 21 | AAV70766 |
| 20 | 153.5 | 3.7 | 526 | 13 | AAR26174 |
| 21 | 150 | 3.6 | 579 | 21 | AAG41949 |
| 22 | 148.5 | 3.6 | 491 | 21 | AAV70770 |
| 23 | 146.5 | 3.5 | 492 | 23 | ABB07836 |
| 24 | 145 | 3.5 | 918 | 22 | AAU30491 |
| 25 | 141.5 | 3.4 | 491 | 21 | AAV70769 |
| 26 | 139 | 3.4 | 496 | 21 | AAV70762 |
| 27 | 138.5 | 3.3 | 416 | 22 | ABB58589 |
| 28 | 138.5 | 3.3 | 645 | 22 | AAG66531 |
| 29 | 138.5 | 3.3 | 646 | 22 | AAU30243 |
| 30 | 138 | 3.3 | 522 | 19 | AAW69726 |
| 31 | 136 | 3.3 | 466 | 23 | ABG60996 |
| 32 | 136 | 3.3 | 470 | 23 | ABG61002 |
| 33 | 136 | 3.3 | 495 | 19 | AAW70270 |
| 34 | 136 | 3.3 | 495 | 23 | ABG61000 |
| 35 | 136 | 3.3 | 1579 | 23 | ABB91088 |
| 36 | 135.5 | 3.3 | 510 | 9 | ABP80615 |
| 37 | 135 | 3.3 | 460 | 21 | AAG41950 |
| 38 | 135 | 3.3 | 472 | 23 | AAG66107 |
| 39 | 135 | 3.3 | 472 | 23 | AAG66113 |
| 40 | 135 | 3.3 | 491 | 13 | AAR25296 |
| 41 | 135 | 3.3 | 497 | 23 | AAG66111 |
| 42 | 135 | 3.3 | 508 | 9 | AAP80664 |
| 43 | 135 | 3.3 | 508 | 13 | AAR25297 |
| 44 | 135 | 3.3 | 508 | 15 | AAR51696 |
| 45 | 135 | 3.3 | 513 | 22 | AAG65635 |

ALIGNMENTS

| | |
|----------|---|
| RESULT 1 | |
| ABG61812 | |
| ID | ABG61812 standard; Protein; 807 AA. |
| AC | ABG61812; |
| DT | 15-AUG-2002 (first entry) |
| DE | Prostate cancer-associated protein #13. |
| KW | Prostate cancer; prostate tumour tissue; human; mammal; cytostatic. |
| OS | Mammalia. |
| XX | |
| PN | WO200230268-A2. |
| PD | 18-APR-2002. |
| XX | |
| PF | 12-OCT-2001; 2001WO-US32045. |
| XX | |
| PR | 13-OCT-2000; 2000US-0687576. |
| PR | 08-DEC-2000; 2000US-0733288. |
| PR | 08-DEC-2000; 2000US-0733742. |
| PR | 24-JAN-2001; 2001US-263957P. |
| PR | 16-MAR-2001; 2001US-276791P. |
| PR | 16-MAR-2001; 2001US-276888P. |
| PR | 06-APR-2001; 2001US-281922P. |
| PR | 24-APR-2001; 2001US-286214P. |
| PR | 30-APR-2001; 2001US-0847046. |
| PR | 04-MAY-2001; 2001US-288589P. |
| XX | |
| PA | (EOSB) EOS BIOTECHNOLOGY INC. |
| XX | |

Human protein disu
Human protein assoc
Human cancer assoc
Human ovarian anti
Human phospholipas
Aspergillus sp. re
Protein disulfide
Protein disulfide
Spider protein dis
Protein disulphide
Arabidopsis thalia
Scorpion protein d
Human protein iden
Novel human secret
Scorpion protein d
Spider protein dis
Drosophila melanog
Human interferon-a
Novel human secret
Saccharomyces cere
Dirofilaria immiti
Dirofilaria immiti
Dirofilaria immiti
Herbicidally activ
Bovine protein thia
Arabidopsis thalia
D. immitis transgl
Recombinant PDI (A
D. immitis transgl
Polypeptide with P
PDI. Homo sapiens
Human PDI. Homo s
Amino acid sequenc

PI Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;
 XX WPI; 2002-471335/50.
 DR N-PSDB; ABR92127.
 XX
 PT Detecting a prostate cancer-associated transcript in a cell in a
 PT patient, useful for diagnosing prostate cancer (PC) or screening
 PT modulators of PC, by determining if prostate cancer-associated genes
 PT are expressed in a prostate tissue -
 XX
 PS Claim 27, Page 311, 436pp: English.
 CC
 CC The present invention relates to methods of detecting a prostate
 CC cancer-associated transcript in a cell from a patient. The method
 CC comprises contacting a biological sample from the patient with
 CC prostate cancer-associated polynucleotides (designated PC genes) that
 CC selectively hybridise to a sequence that is at least 80% identical
 CC to them. The prostate cancer-associated polynucleotide sequences
 CC are differentially expressed in prostate tumour tissue or in
 CC prostate cancer and are derived from the tissues of various
 CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).
 CC The methods of the invention are useful for diagnosing and treating
 CC prostate cancer in mammals. The prostate cancer-associated genes are
 CC useful for diagnosing or treating prostate cancer, as well as for
 CC identifying modulators of prostate cancer or agents that inhibit
 CC prostate cancer. The nucleic acid sequences are particularly useful
 CC in gene therapy, as a vaccine or in antisense applications.
 CC ABB61800-ABB61944 represent prostate cancer-associated proteins.
 CC
 XX
 SQ Sequence 807 AA;
 Query Match 100.0%; Score 4142; DB 23; Length 807;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 807; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 QY 1 MFSGFNFRGISFVICFYMPVNSLPGLSPKYSFTIQPGLEENEARPRQDYGIS 60
 DB 1 MFSGFNFRGISFVICFYMPVNSLPGLSPKYSFTIQPGLEENEARPRQDYGIS 60
 QY 61 VAKNCVKEEISRYCGKEKDKMAYLFKGNLLREFPTDLEFDNNAVIAHFLFALLSEV 120
 DB 61 VAKNCVKEEISRYCGKEKDKMAYLFKGNLLREFPTDLEFDNNAVIAHFLFALLSEV 120
 QY 121 KYTNLEDLONENALKGKANIIFSYVRAIGIPRHRAVMGEVYGTYYQFVLTETALL 180
 DB 121 KYTNLEDLONENALKGKANIIFSYVRAIGIPRHRAVMGEVYGTYYQFVLTETALL 180
 QY 181 ESTGSEVEVAHLYEFCKIVLDTLOCCRTLMROPITTTINHLFITMKRAPLLEVAED 240
 DB 181 ESTGSEVEVAHLYEFCKIVLDTLOCCRTLMROPITTTINHLFITMKRAPLLEVAED 240
 QY 241 POOVSTVHLQGLPLVFIVSQATYEADRTAEVAVWRLLGKAGVLLLRDSLEVNTPOD 300
 DB 241 POOVSTVHLQGLPLVFIVSQATYEADRTAEVAVWRLLGKAGVLLLRDSLEVNTPOD 300
 QY 301 ANVFKRAEGGVPEFVLVDVLLISHVENNMHIEIOEDEDDMEGPDIDVDDEVAE 360
 DB 301 ANVFKRAEGGVPEFVLVDVLLISHVENNMHIEIOEDEDDMEGPDIDVDDEVAE 360
 QY 361 TVPRDRRKRLPLELTVLLEETFNATYMASDSYLFYAGQAVAMALQSTIDVAVKLG 420
 DB 361 TVPRDRRKRLPLELTVLLEETFNATYMASDSYLFYAGQAVAMALQSTIDVAVKLG 420
 QY 421 TSMLLTRINCADMSDYCTQNTVEFPIIKMYKGENPVSYAGMLGTRDKLLFKTQLNRIS 480
 DB 421 TSMLLTRINCADMSDYCTQNTVEFPIIKMYKGENPVSYAGMLGTRDKLLFKTQLNRIS 480
 QY 481 YPVNITSIOGAEEYLGELYKDLIVSSVSVALGFSPMTKANEDFSEAGNYLKGVIYTG 540
 DB 481 YPVNITSIOGAEEYLGELYKDLIVSSVSVALGFSPMTKANEDFSEAGNYLKGVIYTG 540
 QY 541 TYSEEDVLLSTKYAASLPALLARHTEGKIESIPLASTHAQDITVOITTDALLEMPEIT 600
 DB 541 TYSEEDVLLSTKYAASLPALLARHTEGKIESIPLASTHAQDITVOITTDALLEMPEIT 600

DB 541 TYSEEDVLLSTKYAASLPALLARHTEGKIESIPLASTHAQDITVOITTDALLEMPEIT 600
 QY 601 VENLPSYERLOKPLLIIFSDGTVPNOYKKAILLTVKOKYLDSPFCMLNKNTPVGGIL 660
 DB 601 VENLPSYERLOKPLLIIFSDGTVPNOYKKAILLTVKOKYLDSPFCMLNKNTPVGGIL 660
 QY 661 RAYEDPLPPLVLVNLHSGGVFAFPSPDAIIEENLVMLKLEAGLENNHTILPAOE 720
 DB 661 RAYEDPLPPLVLVNLHSGGVFAFPSPDAIIEENLVMLKLEAGLENNHTILPAOE 720
 QY 721 WKPLPAYDPLSMIDATSGRTKRVKCKMETVOENDKEQHBDSKAVKREPJETLRK 780
 DB 721 WKPLPAYDPLSMIDATSGRTKRVKCKMETVOENDKEQHBDSKAVKREPJETLRK 780
 QY 781 HMNRSNMFKEAKESFRRDKEIGSCSVN 807
 DB 781 HMNRSNMFKEAKESFRRDKEIGSCSVN 807
 RESULT 2
 AAB94073
 ID AAB94073 standard; Protein; 357 AA.
 XX
 AC AAB94073;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:14262.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PE 28-JUL-2000; 2000EP-0116126.
 PF
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 14262; 2537bp + CD ROM; English.
 CC
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo- or primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH1628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 357 AA;

Query Match 44.4%; Score 1839; DB 22; Length 357;
 Best Local Similarity 99.7%; Pred. No. 2.6e-155;
 Matches 356; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 451 MYKGENVSYAGMGTDLKFTOLNRSYVNTSIOEAEEYLSGELYKDLITYSSVS 510
 Db 1 MYKGENVSYAGMGTDLKFTOLNRSYVNTSIOEAEEYLSGELYKDLITYSSVS 60
 QY 511 VLGLFSPMTAKEDFESAGNYLKGVTGYTGYSEEDVLLSPKYAASLPALLARHTGK 570
 Db 61 VLGLFSPMTAKEDFESAGNYLKGVTGYTGYSEEDVLLSPKYAASLPALLARHTGK 120
 QY 571 IESIPLASTHAODIVQITTDALLEMFPEITVENLPSYRLOKPLILFSDGTVPNQYKA 630
 Db 121 IESIPLASTHAODIVQITTDALLEMFPEITVENLPSYRLOKPLILFSDGTVPNQYKA 180
 QY 631 ILTVKQKYLDSFTPCWNLKNTKPYRGILRAYFPPLPLVLYVNLHSGQVAFPSD 690
 Db 181 ILTVKQKYLDSFTPCWNLKNTKPYRGILRAYFPPLPLVLYVNLHSGQVAFPSD 240
 QY 691 QAIIENLVMLKLEAGLENNHITLPAQEMKPLPAYDFLSMIDATSORGTTRVPCW 750
 Db 241 QAIIENLVMLKLEAGLENNHITLPAQEMKPLPAYDFLSMIDATSORGTTRVPCW 300
 QY 751 KETDVOENDKQHEKDSAVRKEPIETLRKIHNNRSNMEKAEKSPRRDKELGCSKVN 807
 Db 301 KETDVOENDKQHEKDSAVRKEPIETLRKIHNNRSNMEKAEKSPRRDKELGCSKVN 357

RESULT 3

AAH9829 ID AAR49829 standard; protein; 485 AA.

AC AAR49829;

DT 29-SEP-1994 (first entry)

XX Highly heat-resistant protein disulphide isomerase.

KW Protein disulphide isomerase; PDI; 3-D; heat; resistant;
 catalysis; enzyme; ribonuclease A; substrate; refolding.

OS Humicola insolens KASI.

PN JPO6038752-A.

PD 15-FEB-1994.

PF 04-MAR-1993; 93JP-0044013.

PR 27-MAY-1992; 92JP-0135254.

XX (TOYAMA) TOYAMA CHEM CO LTD.

XX WPI; 1994-094827/12.

PT Novel protein disulphide isomerase active substance - for use in
 the production of 3-D active proteins

PS Claim 2; Page 7-8; 10pp; Japanese.

XX The new PDI has the following characteristics: (A) has an ability of

CC catalysing disulphide exchange reaction in a protein; (B) uses at
 CC least ribonuclease A as the substrate; (C) has an optimum acting
 CC temp. of 20 to 70 degrees C; (D) is stable at pH 6 to 9; and (E) has a
 CC mol. wt. of ca. 60000 to 62000 measured by SDS-PAGE. PDI can be used
 CC advantageously in a refolding reaction of a protein.

XX Sequence 485 AA;

Query Match 4.5%; Score 185; DB 15; Length 485;
 Best Local Similarity 21.9%; Pred. No. 1.9e-07;
 Matches 112; Conservative 79; Mismatches 213; Indels 108; Gaps 22;

QY 15 VIMCIYFMYTVNSLDELSPQKRYSTLQGLLELNAPRLQDYGISVAKNCVKKEISRY 74
 Db 20 LVLAEPFAPWCGCHKALAP-----EYEAATLKKNKIKLAKVDT----- 60
 QY 75 CGKEKDKAKAYLFKGNILREF-----PTTLEDVNAIVAHVLPALL--FSEVKYITN 125
 Db 61 --EETDLCQOHGVEGYPRLKVFRLGIDNVSPLYGQKRAAITSYMIKOSLPAYSEVTK-DN 117
 QY 126 LEDLQNIEMALKKANITFSYVRAIGIPHRVMEAGFYGTQYFVLTTETALLSTGS 185
 Db 118 LEEFKKADKA-----VLVAAYDASDKASSEVFQYAEKLRDNYPPGSSSDAALAEAGV 171
 QY 186 EDVEYAHLYFFHCKLVLDLTQCCRTIMEQPLTTINHLFTMKAPLLTEVAEDPOQVS 245
 Db 172 KAP-----ATVLKDFDEGKAVFSEKFEVAIEKFGATGATPLIGETG--PEYTS 219
 QY 246 TVHQLGLPLRYFVSQA-----TYEADRTAEW--VAMRLIG-KAGVLLLR 290
 Db 220 D-YMSAGIPLAYIFAEETEEERKELSDKLKPIEAORGVINFTIDAKAFGAHAGNINLKT 278
 QY 291 DSLEY-----NIPQDANVVFKRAEGVPYFVLHVDLILSHVENNMHTEQDEEDNDM 346
 Db 279 DKFPAPAIQEVAKNOKFPDQKE--TFEPAIAEVDDFPAKIEPISSEPIEKO---- 333
 QY 347 EGPDDVDVDEVAEFVFRDKRKLPLELTVLETFENATVWASIVLVFYAGQWAVSMA 406
 Db 334 EGPVTVVVAKNNEIVLDDTK-----DYLIEFVAPWCGCHCA 370
 QY 407 FLQSTIDVA--VKLGSTMLLTRINCADMSDVCTKQNTVEFPITIKMKKE--NPVSY 461
 Db 371 LAPKYEELGALYAKSEFDRVIAKVD--ATANDV--PDEIQGFPIKUYPAQAKQAPVY 427
 QY 462 AGMLGTDLKRFTOLNRSYVNTSIOEAE 493
 Db 428 SGRSTVEDLIKFLAENG-KYKAIS--EDAE 456

RESULT 4

AAH73963 ID AAR73963 standard; Protein; 505 AA.

AC AAR73963;

DT 12-JAN-1996 (first entry)

XX Humicola insolens protein disulphide isomerase.

KW Humicola insolens; protein disulphide isomerase; Bacillus brevis;
 promoter; recombinant production; fungal protein.

OS Humicola insolens.

PN JPO7107980-A.

PD 25-APR-1995.

XX Key Location/Qualifiers

XX Peptide 1..20

XX Peptide 21..505

XX Peptide /label-mat_peptide

Db 391 LAPVEELGALYAKSEKDRVIAKVD-ATANDV--PDEIOGFPIIKLYPAGAKGQPTVY 447
QY 462 AGMGTCKDLKFIOLNRIISYPVNTSIOAE 493
Db 448 GSRRTVEDLIKFIENG-KYKAIS--EDAE 476

RESULT 6
AB62855
ID AB62855 standard; Protein; 496 AA.

AC AB62855;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 15357.

KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR N-PSDB; ABL06958.

PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

PS Disclosure; SEQ ID NO 15357; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB116176-AB130511), expressed DNA sequences (AB101840-AB116175) and the encoded proteins (AB57737-AB572072).

CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 496 AA;

Query Match 4.3%; Score 177; DB 22; Length 496;
Best Local Similarity 19.9%; Pred. No. 1e-06;
Matches 99; Conservative 96; Mismatches 180; Indels 122; Gaps 25;

QY 373 ELTVELTEFNATV-----MASDSIVL--FYAGQAVSMFLQSYIDVAVKL-KGST 423

Db 20 EAEEVVEGVLVATVDFKQLADNEFVLEFAPWCGHCKLAPYAKAQQLEKESP 79

QY 424 MLTRINCADSDVCTKONTVEFPIIKMKKGENPVSYAGMLGTDLKFIOLNRIISYPV 483

Db 80 IKLAVADVVEGELAEQAVRGVPLTKFRRSG-SPVEYSGGQADIIAWVTKKTPPAK 138

QY 484 NITSIOAEVYISGELYLDLIYSSVVLGFSPTMKTAKEPSENGNLYKGYITIGIYS 543

Db 139 DLTSVADAEQFL-----KD-----NEIATIGFFKDESEAKTFTKVANAIDSEVF-GVSS 188
QY 544 EEDVLLSTKRYAASLPALLARHTEGKTESIPLASTHAGDIYQIITDALLEMEPEITVEN 603
Db 189 NADVI---AKTEAKDNGVYLFKEFPDK-----KSYFE--GELNEEN 224
QY 604 LPSYFRLQK-PLLIFSDGTVPQY---KKAITLVKQKYLDSFTPCMLNKNTPVVGNG 658
Db 225 LKFAQVQSLPLIVDFNHSASKIFGSIKSHLFFVSE-----GG 266
QY 659 ILRAYDPLPLPL-----LVLVNLHSGQ---VPAF----- 687
Db 267 HIEKYVDPLKEIAKKYRDDILFVTISDEDEHTRIFEEFGMKKEVPYTRILIKLEDMAK 326
QY 688 --PSPQATIEENLVLMKK-LEAGLENHTTI--LPAQEN-KRP---LPAVDLSM-IDAA 737
Db 327 YKPESDDLSAETIEAFLEKFKLDGKLQHLQSELP-EDWDKNPVKVLSSNESVALDS 385
QY 738 TSQGRTRKVP---KCMKETDVOENDEKQEDKSAV-----RKPEIETLRIRKHNRSNW 787
Db 386 KSVLYEFYAPMGCHCKQAPLYDQLAEKYKQMDIVIAKMDSTANLESIKISSEPTIKY 445
QY 788 FKEAEK---SFRDKEL 801
Db 446 FRKEDNKVIDENLDRTL 462

RESULT 7
AAR82801
ID AAR82801 standard; Protein; 505 AA.

AC AAR82801;

DT 29-NOV-1995 (first entry)

DE Bovine phospholipase C-alpha (PLC-alpha).

KW Phospholipase C; PLC; alpha; anti-inflammatory; cancer.

OS Bos taurus.

PN WO9508624-A.

PD 30-MAR-1995.

PF 22-SEP-1994; 94WO-JP01572.

PR 24-SEP-1993; 93JP-0238402.

PA (SHIO) SHIONOGI & CO LTD.

PI Hirai H, Hirano N;

DR WPI; 1995-139593/18.

DR N-PSDB; AAT00952.

PT Human phospholipase C-alpha and DNA encoding it - useful as anti inflammatory

PS Example 1; Page 25-29; 39pp; Japanese.

CC AAR82801 is bovine phospholipase C-alpha (PLC-alpha). PLC-alpha can be used as an anti-inflammatory and may also be used to measure cancer progression.

CC Sequence 505 AA;

Query Match 4.2%; Score 175.5; DB 16; Length 505;
Best Local Similarity 23.3%; Pred. No. 1.4e-06;
Matches 66; Conservative 56; Mismatches 128; Indels 33; Gaps 9;

QY 376 VELTEETFNATVMASDS---IVLFYAGQAVSMFLQSYIDVAVKLKSTMLTRINC 431

Db 28 LETLDNEESRITDTGSSGLMVEFFAPWCCHKIAPYEAAATRLKG--IVPLAKVDC 85
 QY 432 ADMSDVCTKONTEPPIIKMYKGENPVSYAGMLGTDLKFIOLNRISYPVNITSIOEA 491
 Db 86 TANTUMCNKYGSGYPTLKFIFRDGESSAYDGPRTADGIVSHLKKQAGASVPKSEEF 145
 QY 492 EHYLSGELYKDLILYSSVYGLFSPYTKTAKEDFSEAGNYLKGYITTYISEEDVLLS 551
 Db 146 EKFIIDK-----DASVGFPPKDLFSEAHSEFLKASNLRD--NYRFAHTNYESLV 193
 QY 552 TKYASLPALLAR--HREGKIESIPLASTHAQ---DIVQITTDALLMPEPEITVENLP 605
 Db 194 NKYDDDGECITLFRPSHLTNKPEDKTVATEQKMTSGKRIQENITFICHTMEDND 253
 QY 606 SYFRIQ-KPLLIFSD-----GTVPNPKYKAILTLVKQKYLDS 642
 Db 254 --LLOGKDLIAYYDVDEKNAKGSNMYRNRMMVAKKFLDA 293

RESULT 8

AAR69506
 ID AAR69506 standard; Protein; 511 AA.

AC AAR69506;

DT 18-AUG-1995 (first entry)

DE Aspergillus sp. recombinant protein-disulfide-isomerase.

KW Protein-disulfide-isomerase; enzyme; scleroprotein;
 KM filamentous fungus.

XX Aspergillus.

OS Aspergillus.

PN W09500636-A.

PD 05-JAN-1995.

PF 28-JUN-1994; 94WO-DK00266.

PR 28-JUN-1993; 93DK-0000768.

XX (NOVO) NOVO-NORDISK AS.

PA HJort CM;

PI WPI: 1995-052073/07.

DR N-PSDB: AAQ81934.

XX Protein dt-sulphide isomerase enzyme - obtd. from Aspergillus,

PT useful for e.g. treating scleroproteins such as human hair or for

PS fabric cleaning

XX Disclosure; Page 10; 61pp; English.

XX This sequence encodes a protein-disulfide-isomerase (PDI)

CC which is useful for treatment or degradation of scleroproteins,

CC especially hair, skin and wool, etc. The protein is prepared

CC recombinantly from Aspergillus oryzae IFO 4177 or Aspergillus

CC niger A524. The protein may be truncated into various peptides,

CC all with PDI activity. This protein differs from AAR69505 in that

CC AA 511 is Ala not Glu.

XX Sequence 511 AA;

Query March 4.28; Score 172.5; DB 16; Length 511;

Best Local Similarity 19.98; Pred. No. 2.7e-06; Mismatches 181; Gaps 28;

Matches 105; Conservative 107; Indels 131;

QY 15 VIMCFYWPYTNLSPELSPQYFSTLQGLELNEAVRPLQDYGISVAKVNCVKEE-ISR 73

Db 48 LVLAEFAPMGCHKALAP-KY-----EQATLELKKNIPLVAVDCTEEALCR 95

QY 74 YGKEKDLKAY----LFGNILLREPPPTDLPDVAIYAHVLFALLESEVKIT--NLE 127
 Db 96 DQGE-----GPTLKFIFGLDAVK--PYQARQTEAIVSYWKSL-PAVSPVPEMLE 147
 QY 128 DLQNIENALKGKANIIFSVRAIGIPEHRAVAGVYGT-----TQGVLTTEALL 180
 Db 148 EIKTMKT-----VIGTIASD-----QTANDITTFPESQNDYLFANTSASTA 194
 QY 181 ESTGSEDEVAHLYPEHCKLVLDLTOOCCRTLMEDPLTTLNHLFTYKAPLLEVAED 240
 Db 195 KAGVYQP-----SIVLYKDFDEKKATYDGEIEODALLSMWKASTPVGELG-- 242
 QY 241 POOVSTVHLQGLPLVLYSOQATYADRRTA--WVARRLKAGVLLLDLSLEVINP 298
 Db 243 PETYSQ-YITAGIPLAIPA-ETKEERQFTTEEFKIAKHGSGINIVTI----- 290
 QY 299 QDANVYFKAEGVPEFLVDVLDLISHVENMHIE-----EIQDEDN----- 344
 Db 291 -----DAKLYGAHA-GNLMDSKPPAPAIQDPEKNAKYPYDQ 327
 QY 345 --DMEGPDI-----DVODEVAETVFRD--RRKRLPELTVELTEEPNATVMASD 394
 Db 328 SKEVAKADIGKFLQVLDLDDKVEPSIKSEALPTEGEPVTVV--ASHYKDLVDNEKDV 384
 QY 395 L--FTAGQAVSMALQSTIDVAVKLGKSTMLTRINCADMSDVCTKONTEPPIIKMY 452
 Db 385 LLEFYAPMGCHKALAPKEELASLYKDIPEVTIAKID-ATANDV--PDSINGFPPIKLF 441
 QY 453 KKG--ENPYSAGMLGTDLKFIOLNRISYPVNITSIOEAEEYISGE 498
 Db 442 AAGAKDSPVEYEGSRTVEDLANFVKENG-KHKVDALVEDPKKEQESGD 488

RESULT 9

AAR69504
 ID AAR69504 standard; Protein; 515 AA.

AC AAR69504;

DT 18-AUG-1995 (first entry)

DE Aspergillus sp. recombinant protein-disulfide-isomerase.

KW Protein-disulfide-isomerase; enzyme; scleroprotein;
 KM filamentous fungus.

XX Aspergillus.

OS Aspergillus.

PN W09500636-A.

PD 05-JAN-1995.

PF 28-JUN-1994; 94WO-DK00266.

PR 28-JUN-1993; 93DK-0000768.

XX (NOVO) NOVO-NORDISK AS.

PA HJort CM;

PI WPI: 1995-052073/07.

DR N-PSDB: AAQ81933.

XX Protein dt-sulphide isomerase enzyme - obtd. from Aspergillus,

PT useful for e.g. treating scleroproteins such as human hair or for

PS fabric cleaning

XX Disclosure; Page 32; 61pp; English.

XX This sequence encodes a protein-disulfide-isomerase (PDI)

CC which is useful for treatment or degradation of scleroproteins,

CC especially hair, skin and wool, etc. The protein is prepared

CC recombinantly from Aspergillus oryzae IFO 4177 or Aspergillus

CC niger A524. The protein may be truncated into various peptides,
CC all with PDI activity.

XX
XX
SQ Sequence 515 AA;

Query Match 4.2%; Score 172.5; DB 16; Length 515;
Best Local Similarity 19.9%; Pred. No. 2.7e-06;
Matches 105; Conservative 107; Mismatches 185; Indels 131; Gaps 28;

```

QY 15 VIMCIFYMPTVNSLPESLPQKFFSTLPQLENEAVRPLQDYGISAVKNCVKEE-ISR 73
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 48 LVLAFFAPWCGCHKALAP-KY-----EQAAATELKEKNIPLVAVDCTEEBALCR 95

QY 74 YCGKEKDLKAY---LFGKNILLREFPTDLEFVNAYAHVLFALLFSEVKYIT--NLE 127
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 96 DOGVE-----GYPLTKIFRGDLAVK--PYQGARQTEAIVSYMKSL-PAVSPVPENIE 147

QY 128 DLQNIENALKGANIIFSVRAIGIPEHRAVMEAGVYGT-----TYQFVLITTEIALL 180
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 148 EIKTMDKI-----VIGIYASDD-----QTANDIFTFAESQORDNYLFAATSDASTA 194

QY 181 ESIGSEDEYVAHLVYFHCNKLVDLFTQOCRRITMEQPLTTLNHLFIKTKAKALLTEVAD 240
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 195 KAEGVKQP-----SIVLYKDFDEKKATYDGEIEDALLSWKTKASTPLVGELG-- 242

QY 241 PQGVSTVHLQGLPLVFIYSQOATYEADRRPAE--WVAWRLLGKAGVLLLRDSEVNIP 298
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 243 PETYSG-YITAGIPLAYIPA-ETKEERQFTFEFKFAEKHKSINIVYI----- 290

QY 299 QDANVYFKRAEGVPEFLVLDVLLISHVENNMHIE-----EIODEDN----- 344
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 291 -----DAKLYGAHA-GNLLNDPSKPPAFALQDPEKNAKYPYDQ 327

QY 345 --DMEGPDI-----DVODEVAETVFRD---RKRKLPLELTVELTEEFNATVMAVSDSIV 394
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 328 SKEVAKADIGKFIQDVLDDKVEPSIKSEAIPEIQEGPVTVV---AHSYKDLVLNKKDV 384

QY 395 L--FYAGQAVSMAPLQSTIDVAVKIKGSTMLLRINCADMSDVCTKQNTVEFPIKMY 452
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 385 LLEFYAPWCGCHKALAPKYEEELASLYKDIPEVTIAKID-ATANDV--PDSITGFPIKLF 441

QY 453 KKG--ENPVSYAGMLGTQDLKFIQNLRIISYVNTTSIOEAEEYLSGE 498
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 442 AAGAKDSPEYEGSRTVEDLANFVKENG-KHKVDALVDPKKEQESGD 488

```

RESULT 10
AAB49946
ID AAB49946 standard; Protein; 515 AA.

XX
XX
AC AAB49946;

XX
XX
DT 07-MAR-2001 (first entry)

XX
XX
DE A. oryzae protein disulfide isomerase #1.

XX
XX
KW Disulphide bond; allergy; protein disulfide isomerase variant; detergent;
food additive; cosmetic.

XX
XX
OS Aspergillus oryzae.

XX
XX
PN MO200070064-A1.

XX
XX
PD 23-NOV-2000.

XX
XX
PF 17-MAY-2000; 2000WO-DK00265.

XX
XX
PR 17-MAY-1999; 99DK-0000683.

XX
XX
PR 18-MAY-1999; 99DK-0000689.

XX
XX
PR 02-JUN-1999; 99US-0137068.

XX
XX
PA (NOVO) NOVO NORDISK AS.

PI Hjort CM;

XX
XX
DR WPI; 2001-070776/08.

XX
XX
DR N-PSDB; AAC89028.

PT Protein disulfide isomerase variant having increased reducing
PT properties and decreased redox potential than native proteins, used to
PT reduce allergenicity of allergic proteins in feed, food or cosmetic
PT products -

XX
XX
PS Claim 3; Page 72-74; 82pp; English.

CC The present invention provides variants of the Aspergillus oryzae protein
CC disulfide isomerase enzyme. These are capable of reducing disulphide
CC bonds in proteins, which may be the cause of allergies in humans. The
CC proteins can be used to reduce the allergenicity of foods, to treat
CC scleroproteins, in detergents, in food additives and in cosmetics.

XX
XX
SQ Sequence 515 AA;

Query Match 4.2%; Score 172.5; DB 22; Length 515;
Best Local Similarity 19.9%; Pred. No. 2.7e-06;
Matches 105; Conservative 107; Mismatches 185; Indels 131; Gaps 28;

```

QY 15 VIMCIFYMPTVNSLPESLPQKFFSTLPQLENEAVRPLQDYGISAVKNCVKEE-ISR 73
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 48 LVLAFFAPWCGCHKALAP-KY-----EQAAATELKEKNIPLVAVDCTEEBALCR 95

QY 74 YCGKEKDLKAY---LFGKNILLREFPTDLEFVNAYAHVLFALLFSEVKYIT--NLE 127
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 96 DOGVE-----GYPLTKIFRGDLAVK--PYQGARQTEAIVSYMKSL-PAVSPVPENIE 147

QY 128 DLQNIENALKGANIIFSVRAIGIPEHRAVMEAGVYGT-----TYQFVLITTEIALL 180
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 148 EIKTMDKI-----VIGIYASDD-----QTANDIFTFAESQORDNYLFAATSDASTA 194

QY 181 ESIGSEDEYVAHLVYFHCNKLVDLFTQOCRRITMEQPLTTLNHLFIKTKAKALLTEVAD 240
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 195 KAEGVKQP-----SIVLYKDFDEKKATYDGEIEDALLSWKTKASTPLVGELG-- 242

QY 241 PQGVSTVHLQGLPLVFIYSQOATYEADRRPAE--WVAWRLLGKAGVLLLRDSEVNIP 298
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 243 PETYSG-YITAGIPLAYIPA-ETKEERQFTFEFKFAEKHKSINIVYI----- 290

QY 299 QDANVYFKRAEGVPEFLVLDVLLISHVENNMHIE-----EIODEDN----- 344
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 291 -----DAKLYGAHA-GNLLNDPSKPPAFALQDPEKNAKYPYDQ 327

QY 345 --DMEGPDI-----DVODEVAETVFRD---RKRKLPLELTVELTEEFNATVMAVSDSIV 394
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 328 SKEVAKADIGKFIQDVLDDKVEPSIKSEAIPEIQEGPVTVV---AHSYKDLVLNKKDV 384

QY 395 L--FYAGQAVSMAPLQSTIDVAVKIKGSTMLLRINCADMSDVCTKQNTVEFPIKMY 452
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 385 LLEFYAPWCGCHKALAPKYEEELASLYKDIPEVTIAKID-ATANDV--PDSITGFPIKLF 441

QY 453 KKG--ENPVSYAGMLGTQDLKFIQNLRIISYVNTTSIOEAEEYLSGE 498
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 442 AAGAKDSPEYEGSRTVEDLANFVKENG-KHKVDALVDPKKEQESGD 488

```

RESULT 11
AAE12991

XX
XX
ID AAE12991 standard; Protein; 515 AA.

XX
XX
AC AAE12991;

XX
XX
DT 28-JAN-2002 (first entry)

XX
XX
DE Human protein disulphide isomerase (PDI).

XX
XX
KW Heat shock protein; HSP; HSP peptide-binding fragment; HSPF; vaccine;
cytotoxic T cell response; hepatitis virus; herpes simplex virus;

human immunodeficiency virus; bacteria; Mycobacteria; Rickettsia; protozoa; Leishmania; Trypanosoma; intracellular parasite; Chlamydia; sarcoma; carcinoma; cancer; human; protein disulphide isomerase; PDI.

OS Homo sapiens.

XX XX

PH Key Location/Qualifiers

FT Binding-site 1..370

FT Binding-site /note= "Peptide binding domain"

FT Binding-site 204..491

FT Binding-site /note= "Peptide binding domain"

FT Binding-site 213..351

FT Binding-site /note= "Peptide binding domain"

XX XX

PN US2001034042-A1.

XX XX

PD 25-OCT-2001.

XX XX

PF 12-JAN-2001; 2001US-0759010.

XX XX

PR 20-JAN-2000; 2000US-0488393.

XX XX

PA (SRIV/) SRIVASTAVA P K.

XX XX

PI Srivastava PK:

XX XX

DR WPI: 2001-656559/75.

XX XX

PT Vaccine compositions for vaccinating against cancers and infections,

PT comprises peptide-binding fragments (PBFS) of heat shock proteins

PT (HSPs) and non-covalent complexes of PBFS of HSPs and antigenic

PT molecules -

XX XX

PS Disclosure; Fig 3; 39p; English.

XX XX

CC The invention relates to pharmaceutical compositions comprising

CC peptide binding fragments of heat shock proteins (HSPs) and non-

CC covalent complexes of HSP peptide-binding fragments (HBF) in non-

CC covalent association with antigenic molecules. Vaccines comprising

CC peptide fragments of the invention may be used to stimulate an

CC immune response, in particular cytotoxic T cell responses against

CC cells infected with viruses (including hepatitis type A, B and C,

CC influenza, varicella, adenovirus, herpes simplex (HSV) type 1 and

CC type II, rinderpest rhinovirus, echovirus, rotavirus, respiratory

CC syncytial virus, mumps virus, papova virus, papilloma virus,

CC arbovirus, cytomegalovirus, echinovirus, hantavirus, coxsackie virus,

CC measles virus, rubella virus, polio virus, HIV-I, and HIV-II;

CC bacteria including (including Mycobacteria, Rickettsia, Mycoplasma,

CC Neisseria and Legionella); protozoa (including Leishmani, Kokzidia

CC and Trypanosoma) and intracellular parasites (including Chlamydia and

CC Rickettsia). The vaccines may be used to treat cancers such as human

CC sarcomas and carcinomas, pancreatic cancer, breast cancer, ovarian

CC cancer, prostate cancer, squamous cell carcinoma, basal cell carcinoma,

CC adenocarcinoma, sweat gland carcinoma, sebaceous gland carcinoma,

CC papillary carcinoma, papillary adenocarcinomas, medullary carcinoma,

CC cystadenocarcinoma, bronchogenic carcinoma, renal cell carcinoma,

CC hepatoma, bile duct carcinoma, choriocarcinoma, seminoma and embryonal

CC carcinoma. The present sequence is human protein disulphide isomerase

CC (PDI), a HSP involved in folding of many disulphide-bonded proteins.

XX XX

SQ Sequence 515 AA:

Query Match 4.1%; Score 169.5; DB 22; Length 515;

Best Local Similarity 20.6%; Pred. No. 5, 1e-06;

Matches 11; Conservative 89; Mismatches 193; Indels 145; Gaps 27;

OY 15 VMCIPTVPTVNSLPFLSPKYSTSTLOPGLEENAVRPLODYGISVAKYCKEELSRK 74

DB 44 LVLAEEFAWPCGCKALAP-KY-----EEAATELKANNIPDVKDCFAEE----- 87

OY 75 CGEKKLMKAVLEKGNILREF-----PTDTLPVNAIVAVHLPALL--FSEVYVYTN 125

DB 88 -----DLCHSQGVEGYPTLKIFRGVDSKPYQCARQTESIVYMKQSLPVAHSVNE-EN 141

OY 126 LEDLQNIENALKKGNIFSYVAIGIPEHRAVMEAGFYGTGYFLVTEALLLESIGS 185

DB 142 LEEIKTMDKI-----VVGIVIPSDQETVQAREKVAESQRQNYLPAATDAAIAKSEGV 195

OY 186 EDVEYAHLY--FFHCKLVLD--LTQCCRTLMQPLTLNIHLFITMKAPLLEVAEDP 241

DB 196 EDPSTV-LYKDFEKKAVYDGEIEQA-----IHSWVSASTPLVGEIG--P 239

OY 242 QGVSTVHLOGLPVFTVISOQATEYEDRPTAEW-----YAMRLLG-KAGV 285

DB 240 EYISG-YIGAGVPLATIFA-ETKEEREKYTEDEKPAQKHKGAINATIDAKHFGAHAGN 297

OY 286 LLLRLDSLEVINIPD-----ANVVEKRAEE--GVVPEFLVLDVLLISHVENMHIEEQ 339

DB 298 LNLDSQKFPAPAFIQDPKAKNAKPYDQAKELMADEVEKEFI---QDVLDGKVEPSIKSEPV 354

OY 340 EDEDNDMEGPDIDVDQDEVAEYFRDRKRPLPLELTVELTEEFNKTVAASDSIVL--TY 397

DB 355 ESQ-----EGP-----VTV-VVAHSYKDLVIDNDKVLLEFY 385

OY 398 AGQAVSMAFLQSYIDVAVAKLKGSTMLLTRINCADMSDVCYK-----QNV 443

DB 386 APWCGCKALAPKYDELAALY-----ADHPDLAAVYTIKIDATANDVPPDI 432

OY 444 TEPPILKMYKKG--ENPVSYAGMLGTRDLKFTQLNRISYPVNTSIOEAEEYLSGEL 499

DB 433 TGEPTRLYPAGAKSDPIEYSGSRVEDLANFVKENG-KHNVDALNVASBEQEGGDV 489

RESULT 12

AAB82539

ID AAB82539 standard; Protein; 515 AA.

XX XX

AC AAB82539;

XX XX

DT 17-SEP-2001 (first entry)

XX XX

DE Human protein disulphide isomerase.

XX XX

KW Protein disulphide isomerase; PDI; human; heat shock protein; HSP;

KW immunotherapy; therapy; cancer; infection; vaccine.

XX XX

OS Homo sapiens.

XX XX

PH Key Location/Qualifiers

FT Binding-site 1..370

FT Binding-site /note= "peptide-binding domain"

FT Binding-site 5..232

FT Binding-site /note= "peptide-binding domain"

FT Binding-site 213..351

FT Binding-site /note= "peptide-binding domain"

FT Binding-site 204..491

FT Binding-site /note= "peptide-binding domain"

XX XX

PN WO200152791-A2.

XX XX

PD 26-JUL-2001.

XX XX

PF 18-JAN-2001; 2001WO-US01781.

XX XX

PR 20-JAN-2000; 2000US-0488393.

XX XX

PA (UYCO-) UNIV CONNECTICUT HEALTH CENT.

XX XX

PI Srivastava PK:

XX XX

DR WPI: 2001-457506/49.

XX XX

PT Pharmaceutical composition, used to treat or prevent infection or

PT cancer, comprises a complex comprising a heat shock protein-binding

PT fragment associated with a molecule displaying antigenicity of an

PT infectious agent or cancer cell -

XX Claim 53; Fig 3; 106pp; English.

CC The present sequence is that of human protein disulfide isomerase
CC (PDI), a heat shock protein (HSP) involved in the folding of
CC disulfide-bonded proteins. PDI contains various domains that
CC contribute to peptide binding, including a fragment that includes
CC amino acids 213-351 and is capable of binding short (10-15 amino
CC acid) peptides, and a larger fragment, including amino acids
CC 204-491, which is able to bind polypeptides of 25-50 amino acids.
CC An even larger fragment, including amino acids 1-370, is required
CC for more complex substrates, such as larger polypeptides. The
CC invention relates to complexes of peptide-binding fragments of HSPs
CC with antigenic molecules and their use in immunotherapy for the
CC treatment of infectious diseases and cancer. Claimed methods of
CC treating or preventing cancer/infectious disease involve culturing
CC a cancer cell/infected cell transformed with a nucleic acid
CC encoding a HSP peptide-binding domain, recovering complexes of the
CC HSP fragments noncovalently associated with peptides from the
CC cancer cell/infected cell, and administering the recovered
CC complexes. These methods can use PDI peptide-binding fragments
CC comprising amino acids 5-232, 213-351 and 204-491 of the present
CC sequence.

XX Sequence 515 AA;

Query Match 4.1%; Score 169.5; DB 22; Length 515;
Best Local Similarity 20.6%; Pred. No. 5.1e-06;
Matches 111; Conservative 89; Mismatches 193; Indels 145; Gaps 27;

QY 15 VIMCIFYWPTVNSLPDLPSPKRYFTLOPGLLELNEAVRPLDYGISVAKVCKEISRY 74
DB 44 LVLAFFAPWCGCHKALAP-KY-----EEATELKAKNIPLVKVDCTAE----- 87
QY 75 CGKEKDLKATLVFKNILREF-----PTDTLFDNAIYAHVLPALL--FSEVKYITN 125
DB 88 -----DLCRSGVGEGYPLTKIFRGVDSKPYQAGAROTESIYSYMKOSILPAVSSVNE-EN 141
QY 126 LEDLONTENALGKKNANIFSVRAIGPEHRAVMAGFVYGTQYFVLTTETALLSESGS 185
DB 142 LEEKTKMDKT-----VIGTIPSDDETQYAFKEKYSQORDNYPFAATDAAIAKSGEV 195
QY 186 EDVEYAHLY--FFHCKLVLD--LTQOCRRITMEQPLTTLNHLFKTKAKAPLTVEADP 241
DB 196 EGPSTV-LYKDFDEKKAVYDGEIAGEA-----IHSWKASSTPLVGEIG--P 239
QY 242 QGVSTVHLQGLPLVFIYSQATTEADRRITAEW-----VAMRLIG-KAGV 285
DB 240 ETVSG-YIGAGVPLAIFYA-ETKEEREKYTEDEKPIAQKHKGAINIATIDAKMFGAHAGN 297
QY 286 LLLRLDSLEVINIPQD---ANVVERKRAEE--GVPEPIVLHDVDLIISHVENNMHIEIQ 339
DB 298 LNLDSQKPPAPAIQDPKANNAYPPDQAKELNADEVEKFI---QVDLGRKVEPSIKSEPPV 354
QY 340 EDEDNDMEGPDIDVQDEVAETVFRDRKRLPLELTVELTEETFNATYMASDSIYL--FY 397
DB 355 ESQ-----EGP-----VTV--VVAHSYKDYIINDKDYLLFEFY 385
QY 398 AGWQAVSMALQSYIYDAVAKLKGSTMLLTRINCADMSDVCTK-----ONV 443
DB 386 APWCGHCKALAPKYDELALALY-----ADHPDLAKVYIAKIDATFANDVPPPI 432
QY 444 TEFPIIKMYKKG--ENPVSYAGMLCTKDLKFIQNLNRISTYVNTISIOAEETYSGL 499
DB 433 TGFPTLRLRYPGAKDPSLEYSGSRVEDLANFVKENG--KHNVADALNVAASEETOEGDV 489

RESULT 13

AAB43424

ID AAB43424 standard; Protein: 544 AA.

XX AC AAB43424;
XX

DF 08-FEB-2001 (first entry)

XX Human cancer associated protein sequence SEQ ID NO:869.

KW Human: cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnery; immunomodulator;
KW antidiabetic; antisthmatic; antirheumatic; antiarthritic; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening.

OS Homo sapiens.

PN WO20005350-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05882.

PR 12-MAR-1999; 99US-0124270.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

DR WP1; 2000-587533/55.

DR N-PSDB; AAC77633.

PT Novel isolated nucleic acids comprising sequences encoding peptides

PS useful for treating or diagnosing e.g. cancer -

XX Claim 11; Page 1424-1426; 2352pp; English.

CC AACT7607 to AACT78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnery; immunomodulator;
CC antidiabetic; antisthmatic; antirheumatic; antiarthritic;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC agonists or viral infections. The peptides, nucleotides, antibodies,
CC and antagonists may be also be used in drug screens. AACT78449 to
CC AACT78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.

XX Sequence 544 AA;

Query Match 4.1%; Score 169.5; DB 21; Length 544;
Best Local Similarity 22.6%; Pred. No. 5.6e-06;
Matches 64; Conservative 59; Mismatches 127; Indels 33; Gaps 10;

QY 376 VELTEETFNATYMASDS-----IVLFYAGWQAVSMALQSYIYDAVAKLKGSTMLLTRINC 431

DB 67 LEITDDNFESRISDQSGAGMLIVFERFAPWCGCHKRLAPLEYEAARLKG--IVPLAKDC 124

QY 432 ADMSDVCTKQVTEPIIKMYKKGEPVSYAGMLCTKDLKFIQNLNRISTYVNTISIOEA 491
DB 125 TANTWTCNKYGVSGYPLTKIFRDEGEAGAYDGPRTADIVS--HLKKQAGPASVYLPREE 182


```
XX WPI; 1995-139593/18.
DR N-PSDB; AAQ86642.
XX
XX Human phospholipase C-alpha and DNA encoding it - useful as anti
PT inflammatory
XX
XX Claim 1; Page 21-24; 39pp; Japanese.
XX
XX AAR70329 is human phospholipase C-alpha (PLC-alpha). PLC-alpha
CC is useful as an anti-inflammatory and may also be used to measure
CC cancer progression.
XX
SQ Sequence 505 AA;

Query Match 4.0%; Score 166.5; DB 16; Length 505;
Best Local Similarity 22.6%; Pred. No. 9.1e-06;
Matches 64; Conservative 59; Mismatches 127; Indels 33; Gaps 10;

QY 376 VELTEEFENATVMA SDS---IVLFYAGQAVSMATLSQSYIDVAVKLGTSMTLRLINC 431
DB 28 LELTDDNFESRISDTGSAGMLVEFFAPWCGCHKRLAPEYEAARLKG--IVPLAKYDC 85
QY 432 ADMSDVCTKQNVTEPPIIKMYKKGENPVSYAGMLGTDLKFLIQLNRTSYPVNITSIQEA 491
DB 86 TANTNTCNKYGSGVPTLKIFRDGEAGAYDGPRTADGIVS--HLKKQAGPASYPLRTEE 143
QY 492 EEYLSGELYKDLIYSSVSVLGSPMTKTAKEDEFSEAGNYLKGIVTGIYSEEDVLLS 551
DB 144 E-----FKRFISDKDASIVGFDDSFSEAHSEFLKAASNLRD--NYRFAHTNVESTIV 193
QY 552 TKYAASLPALLAR--HTEGKTESIPLASTHAQ---DIVOIITDALLEMPPEITVENLP 605
DB 194 NEYDNGEGIILFRPSHLTNKFYEYKTVAYTEQKMTSGKIKKFIQENIFGICPHMTEDNKD 253
QY 606 SYFRLQ-KPLLILFSD-----GVNPOYKKAILLVLVKQKYLDS 642
DB 254 ---LIQKDLILAYYDVYDEKDAKGSNYWRNRVMVAKKFLDA 293
```

Search completed: March 13, 2003, 17:12:58
Job time : 45 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:12:17 ; Search time 18 Seconds
(without alignments)
1319.128 Million cell updates/sec

Title: US-09-847-046-2

Perfect score: 4142

Sequence: 1 MFSGFNVFRVIGSIFVIMCIF.....FKFEAKSFRRDKELGCKSVN 807

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PT05.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------------|-------------------|
| 1 | 185 | 4.5 | 485 | US-08-068-395A-1 | Sequence 1, Appl |
| 2 | 185 | 4.5 | 485 | US-08-464-365-1 | Sequence 1, Appl |
| 3 | 185 | 4.5 | 505 | US-08-068-395A-3 | Sequence 3, Appl |
| 4 | 185 | 4.5 | 505 | US-08-464-365-3 | Sequence 3, Appl |
| 5 | 175.5 | 4.2 | 505 | US-08-627-907A-4 | Sequence 4, Appl |
| 6 | 172.5 | 4.2 | 511 | US-08-557-122A-4 | Sequence 4, Appl |
| 7 | 172.5 | 4.2 | 511 | US-08-262-666-4 | Sequence 4, Appl |
| 8 | 172.5 | 4.2 | 515 | US-08-557-122A-3 | Sequence 3, Appl |
| 9 | 172.5 | 4.2 | 515 | US-08-557-122A-34 | Sequence 3, Appl |
| 10 | 172.5 | 4.2 | 515 | US-09-262-666-3 | Sequence 3, Appl |
| 11 | 172.5 | 4.2 | 515 | US-09-262-666-34 | Sequence 34, Appl |
| 12 | 166.5 | 4.0 | 505 | US-08-627-907A-2 | Sequence 2, Appl |
| 13 | 152.5 | 3.7 | 530 | US-08-557-122A-35 | Sequence 35, Appl |
| 14 | 152.5 | 3.7 | 530 | US-09-262-666-35 | Sequence 35, Appl |
| 15 | 150 | 3.6 | 521 | US-08-557-122A-32 | Sequence 32, Appl |
| 16 | 150 | 3.6 | 521 | US-09-262-666-32 | Sequence 32, Appl |
| 17 | 138 | 3.3 | 522 | US-09-368-588-2 | Sequence 2, Appl |
| 18 | 137 | 3.3 | 509 | US-08-557-122A-29 | Sequence 29, Appl |
| 19 | 137 | 3.3 | 509 | US-08-262-666-29 | Sequence 29, Appl |
| 20 | 136 | 3.3 | 466 | US-08-984-919A-33 | Sequence 33, Appl |
| 21 | 136 | 3.3 | 470 | US-08-984-919A-55 | Sequence 55, Appl |
| 22 | 136 | 3.3 | 495 | US-08-984-919A-47 | Sequence 47, Appl |
| 23 | 136 | 3.3 | 508 | US-08-557-122A-37 | Sequence 37, Appl |
| 24 | 136 | 3.3 | 510 | US-09-262-666-37 | Sequence 37, Appl |
| 25 | 136 | 3.3 | 510 | US-08-441-139-20 | Sequence 20, Appl |
| 26 | 136 | 3.3 | 510 | US-08-557-122A-30 | Sequence 30, Appl |
| 27 | 136 | 3.3 | 510 | US-09-262-666-30 | Sequence 30, Appl |

| | | | | | |
|----|-------|-----|-----|---------------------|-------------------|
| 28 | 135.5 | 3.3 | 509 | US-08-557-122A-27 | Sequence 27, Appl |
| 29 | 135.5 | 3.3 | 509 | US-09-262-666-27 | Sequence 27, Appl |
| 30 | 135 | 3.3 | 468 | US-08-874-102-33 | Sequence 33, Appl |
| 31 | 135 | 3.3 | 472 | US-08-874-102-55 | Sequence 55, Appl |
| 32 | 135 | 3.3 | 497 | US-07-872-67B-3 | Sequence 3, Appl |
| 33 | 135 | 3.3 | 497 | US-08-874-102-47 | Sequence 47, Appl |
| 34 | 135 | 3.3 | 510 | US-08-557-122A-28 | Sequence 28, Appl |
| 35 | 135 | 3.3 | 510 | US-09-262-666-28 | Sequence 28, Appl |
| 36 | 133.5 | 3.2 | 638 | US-08-557-122A-38 | Sequence 38, Appl |
| 37 | 133.5 | 3.2 | 638 | US-09-262-666-38 | Sequence 38, Appl |
| 38 | 133 | 3.2 | 870 | US-09-134-001C-4959 | Sequence 4959, Ap |
| 39 | 132 | 3.2 | 510 | US-08-557-122A-36 | Sequence 36, Appl |
| 40 | 132 | 3.2 | 510 | US-09-262-666-36 | Sequence 36, Appl |
| 41 | 131.5 | 3.2 | 364 | US-08-650-275-4 | Sequence 4, Appl |
| 42 | 131.5 | 3.2 | 364 | US-09-181-318-4 | Sequence 4, Appl |
| 43 | 130 | 3.1 | 531 | US-08-923-536A-12 | Sequence 12, Appl |
| 44 | 129.5 | 3.1 | 281 | US-08-557-122A-5 | Sequence 5, Appl |
| 45 | 129.5 | 3.1 | 281 | US-09-262-666-5 | Sequence 5, Appl |

ALIGNMENTS

RESULT 1
US-08-068-395A-1
: Sequence 1, Application US/08068395A
: Patent No. 5496719
: GENERAL INFORMATION:
: APPLICANT: YAMADA, YUKIO
: APPLICANT: ASAMI, OSAMU
: APPLICANT: SUGIYAMA, HIDEHIKO
: APPLICANT: IDEKOBAY, CHIE
: APPLICANT: HOSHINO, RUMIHIKO
: APPLICANT: HIRAI, MASANA
: APPLICANT: KAJINO, TSUTOMU
: APPLICANT: IMABEDA, TAKAO
: APPLICANT: SARAI, KIYOKO
: TITLE OF INVENTION: POLYPEPTIDE POSSESSING PROTEIN DISULFIDE
: TITLE OF INVENTION: ISOMERASE ACTIVITY GENE ENCODING THE SAME AND PROCESS FOR
: TITLE OF INVENTION: PRODUCING THE SAME
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
: ADDRESSEE: P.C.
: STREET: 1755 S. Jefferson Davis Highway, Suite 400
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/068,395A
: FILING DATE: 19930527
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 4-135254
: FILING DATE: 27-MAY-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 5-44013
: FILING DATE: 04-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 5-44014
: FILING DATE: 04-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Ohion, No. 5496719man F.
: REGISTRATION NUMBER: 24,618
: REFERENCE/DOCKET NUMBER: 68-228-0
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 413-3000

```

: TELEFAX: (703) 413-2220
:
: TELEX: 248855 OPAT UR
:
: INFORMATION FOR SEQ ID NO: 1:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 485 amino acids
:
: TYPE: amino acid
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
US-08-068-395A-1

```

| | | | | |
|-----------------------|------------------|-------------------|-------------|-------------|
| Query Match | 4.5% | Score 185; | DB 1; | Length 485; |
| Best Local Similarity | 21.9% | Pred No. 1.7e-08; | | |
| Matches 112; | Conservative 79; | Mismatches 213; | Indels 108; | Gaps 22 |

| | | | |
|----|-----|---|------|
| QY | 15 | VIMOFIYFMPYTNLSPELSPKCFESTLOPGBLELNEAVRLODDYSISAKVACVVEEISRY | 74 |
| Db | 20 | LVLAEEFAPWCGHCKALAP-----EYEEAFTTLKEKNIKILAKVDC----- | 60 |
| QY | 75 | CGKEKDLKATLFEKNILIREF-----PDTLTEDYNAIYAHVLFALL--FSEVKTITN | 1255 |
| Db | 61 | --EETDLOQGVGVGYPTLKFRLGDLNVSPPKGGKAAATISYMIKOSLPAVSEVTK-DN | 117 |
| QY | 126 | LEDLONIENTALAKGRANIIFFSVYRAGIPEHNAVMAEGFVYGTTOFQVULTEIALLIESIGS | 1859 |
| Db | 118 | LEEEKKADKA-----VLVAAYVDSKRASSVFLFOVAKLEKLRDNPFSSSDMALAEAGV | 177 |
| QY | 186 | EDVEYAHLYFHCKLVLDLTQOCRRITMEODLTTLNTHLFTKTKAPULTEVAEDPOOVS | 245 |
| Db | 172 | KAP-----AIYLYKDFDECKAFAVSEKFEVEALEKFAKTCATPLIGIG--PETYS | 219 |
| QY | 246 | TVHLTOLGLPVEFIYSOOA-----TYEADRRTAEV--VAMRLTG-KAGVILLLR | 250 |
| Db | 220 | D-VYSAIGIPLAYITAEVRAEEKEKLESDKLKPLAEQKGVINFGTIDAKAFAGAHAGNLUKT | 278 |
| QY | 291 | DSLEY---NIPODANVFKRAEEGVPEVFLVLDHVDLLIISHVENNMHIEIOEDENDM | 346 |
| Db | 279 | DKFPAFAIOEAVKAKKQKPEFDQEK--ITFEFAIKAFVDDFVGAKIEPSTIKSEPIDEKQ--- | 333 |
| QY | 347 | EGPIDVODDEVAETVEPRDRKRKLPLELTYELTBEFTNATVYMASDSVLVEYAGQOANVMA | 406 |
| Db | 334 | EGPITVYVAKNYNEITVLDTK-----DVILEFTAPCGHCCKA | 3707 |
| QY | 407 | FLQSYIDVA--VKLGSTIMLTLTRINCADMSDYCTKQNTYEPFIIMYKRG--NPNVS | 461 |
| Db | 371 | LAPYEEELGALYASSEKDRVIAKVD-ATANDV--DDEIOGFPTIKLYPAGAGAGGVTY | 427 |
| QY | 462 | AGMGTGDKLLKFIOLNRIISYPVNITSIOQAAE | 493 |
| Db | 428 | SGSRTVEDLKIFIAENG-KTKAALS--EDDEE | 456 |

RESULT 2
 US-08-464-365-1
 : Sequence 1, Application US/08464365
 : Patent No. 5700659
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: YAMADA, YUKIO
 : APPLICANT: ASAMI, OSAMU
 : APPLICANT: SUGIYAMA, HIDEHIKO
 : APPLICANT: IDEKUBA, CHIE
 : APPLICANT: HOSHINO, FUMHIKO
 : APPLICANT: HIRAI, MASANA
 : APPLICANT: KAJINO, TSUTOMU
 : APPLICANT: IMAEDA, TAKAO
 : APPLICANT: SARAI, KIYOKO
 :
 : TITLE OF INVENTION: POLYPEPTIDE POSSESSING PROTEIN DISULFIDE
 :
 : TITLE OF INVENTION: ISOMERASE ACTIVITY GENE ENCODING THE SAME AND PROCESS FOR
 :
 : TITLE OF INVENTION: PRODUCING THE SAME
 :
 : NUMBER OF SEQUENCES: 10
 :
 : CORRESPONDENCE ADDRESS:
 :
 : ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
 :
 : ADDRESSEE: P.C.
 :
 : STREET: 1755 S. Jefferson Davis Highway, Suite 400

```

1 CITY: Arlington
2 STATE: Virginia
3 COUNTRY: U.S.A.
4 ZIP: 22202
5
6 COMPUTER READABLE FORM:
7 MEDIUM TYPE: Floppy disk
8 COMPUTER: IBM PC compatible
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10 SOFTWARE: PatentIn Release #1.0, Version #1.25
11
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/08/464,365
14 FILING DATE:
15 CLASSIFICATION: 435
16
17 PRIOR APPLICATION DATA:
18 APPLICATION NUMBER: JP 4-135254
19 FILING DATE: 27-MAY-1992
20
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: JP 5-44013
23 FILING DATE: 04-MAR-1993
24
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: JP 5-44014
27 FILING DATE: 04-MAR-1993
28
29 ATTORNEY/AGENT INFORMATION:
30 NAME: Oblon, No. 5700659man F.
31 REGISTRATION NUMBER: 24,618
32 REFERENCE/DOCKET NUMBER: 68-228-0
33
34 TELECOMMUNICATION INFORMATION:
35 TELEPHONE: (703) 413-3000
36 TELEFAX: (703) 413-2220
37
38 TELEX: 248855 OPAT UR
39
40 INFORMATION FOR SEQ ID NO: 1:
41 SEQUENCE CHARACTERISTICS:
42 LENGTH: 485 amino acids
43 TYPE: amino acid
44 TOPOLOGY: linear
45
46 MOLECULE TYPE: protein
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
96
```

| | | | | |
|---------------------------|--------|-------------------|-------------|-------------|
| Query Match | 4.5%; | Score 185; | DB 1; | Length 485; |
| Best Local Similarity | 21.9%; | Pred No. 1.7e-08; | | |
| Matches 112; Conservative | 79; | Mismatches 213; | Indels 108; | Gaps 22 |

[illegible]

APPLICATION NUMBER: JP 5-44013
FILING DATE: 04-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-44014
FILING DATE: 04-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: OBILO, NO. 570065man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 68-228-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-464-365-3

Query Match 4.5%; Score 185; DB 1; Length 505;
Best Local Similarity 21.9%; Pred. No. 1,9e-08;
Matches 112; Conservative 79; Mismatches 213; Indels 108; Gaps 22;

15 VMCIPTVMTVNSLPKSPKFFSTLQPGLEINEAVRPLODYGISAVAKVCKEISRY 74
40 LVLAEPFAWCGCHKALP-----EYEEAATLTKERIKLAKVCT----- 80
75 CGKENDMAYLEFGNILLREF-----PTDLEFVNAIVANVLEALL--PSEVYITN 125
81 --ETDLCQOHGEGYPTLKVFGRGLDNVSPYKQKRAAITSYMIQSLPAYSEVTK-DN 137
126 LEDLQNIENALGKANIIFSVRAIGIPERRAVMEAGFYGTGYOPVLTEILLESIGS 185
138 LEEFKKADK-----VLVAAYDASDKASSEVFTQYAEKLRDNYPPRGSSDALAEQV 191
186 EDVEYAHLYFFCKLVLDLTQOCRTLMBOPLTTLNHLFIKTKAPLLEVAEDPOQVS 245
192 KAP-----AIVLYDFDEGKAVFSEKEVEVAIEKFAKTGATPLIGEIG--PEYVS 239
246 TVHLDGLPLVIFSOA-----TYEADRTNEM--VANRLIG-KAGVLLLR 290
240 D-YMSGIFLATIFETAEERKELSKLPFAEQGVNFGTIDAKAGAHGNLNLKT 298
291 DSLV-----NIPODANVVERAEQVPELVLDVLIISHVENMHIETIOEDENDM 346
299 DKPFAIQEVAKNOKFPEDQKE-ITFEAKAFVDDFVAGKIEPSIKSEPIEQ----- 353
347 EEPDIDVODEVAEYFRRKRKLPLELTVELTEETFNATWASDSIVLFYAGWQAVSMA 406
354 EGPVTVVAKNYNEIYLDOTK-----DVLIEFYAWCGCHCKA 390
407 FLOSYIDVA---VKLGSTMLLTRINCADMSVCTKQNVTEPILIKMKKG- NPVS 461
391 LAPKEELGALYAKSEFKRVVIAYKD--PDEIQGPTIKLYPAGAKGQPVY 447
462 AGMLGTDLKLFQILNRISYPVNITSIOEAE 493
448 SGRSTVEDIKRIFAENG-KYKAIS--EDAE 476

RESULT 5
US-08-627-907A-4
Sequence 4, Application US/08627907A
Patent No. 6060302
GENERAL INFORMATION:
APPLICANT: HIRANO, Naoto
TITLE OF INVENTION: HUMAN PHOSPHOLIPASE C-ALPHA AND DNA
TITLE OF INVENTION: SEQUENCE ENCODING THE SAME
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:

ADDRESSEE: C/O FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/627,907A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-238402
FILING DATE: 24-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP94/01572
FILING DATE: 22-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: HALEY Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: SHGN-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-627-907A-4

Query Match 4.2%; Score 175.5; DB 3; Length 505;
Best Local Similarity 23.3%; Pred. No. 1.4e-07;
Matches 66; Conservative 56; Mismatches 128; Indels 33; Gaps 9;

376 VELTEETFNATWASDS---YLFYAGWQAVSMAPLOSIVDAVVKLGSTMLTRINC 431
28 LELTDNFSRIITDISSGLMVEFPAPCGCHKRLAPEYEAATRLKG--IYPLAKVVC 85
432 ADMSDVCTKQNVTEPFIIMYKGENPVSYAGMLGTDLKFLQILNRISYPVNITSIOEA 491
86 TANTNTCNKRYGVSYPYTLKIFRQGESGAYDGPRTADGIVSHUKQAGPASYPLKSEEF 145
492 EBYLSELYKDLILYSSVYGLFSPYTKAKEDPSEAGNYLKGYIITGIYSEEDVLLLS 551
146 EKFIISK-----DASVGEFKDLFSEAHSEFLKASNLRD--NYRFAHTNYESLV 193
552 TRYASLPAALLAR--HTGKTESIPLASTHAO---DIVQITDALLMPPEIYENLP 605
194 NKYDDGEGITLFRSHLTNKFEDKTVATTEOKMISGKIKRIFOENIFGICPHMTEDNDK 253
606 SYFRLO-KPILLIFSD-----GTVNPQYKKAILLTVKQYLLS 642
254 ---LLOGKDLLLAYVDVDEKNAKGSNRYNRVMYVAKKFLDA 293

RESULT 6
US-08-557-122A-4
Sequence 4, Application US/08557122A
Patent No. 5879664
GENERAL INFORMATION:
APPLICANT: Hjort, Carsten Malland
TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESS: No. 5879664 No. 5879664disk of No. 5879664th America, Inc.
STREET: 405 Lexington Avenue, 64th floor
CITY: New York

STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,122A
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3980.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 511 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-557-122A-4

Query Match 4.2%; Score 172.5; DB 2; Length 511;

Best Local Similarity 19.9%; Pred. No. 2.7e-07; Matches 105; Conservative 107; Mismatches 185; Indels 131; Gaps 28;

QY 15 VMICFYMPVNSLPESLPQKFFSTLQGLNELNVAVPLDYGISAVKVCKEE-ISR 73
DB 48 LVLAFFAPWCGHCKALAP-KY-----EQAAELKEKNIPLVKVCTEEALCR 95
QY 74 YCGEKDLMKAY----LFKGNILREPTDTLFDVNAIVAHVLAFLSEKYYIT--NLE 127
DB 96 DQGEV-----GYPLKIFRGDAVK--PYGARGTEAIVSVWKOST--PAVSPTPENLE 147
QY 128 DLQIENNAKGNIIISYVAIGIPEHRAVMEAGFYGT-----TYQVLTTEIALL 180
DB 148 EIKTMDTI-----VIGIYASDD-----QTANDITTTAESORDNYLEAATSDASIA 194
QY 181 ESIGSEDEVANHLFFHCKLVLDLTOCCRTLMEOPLTTNLINHLFIKTMKAPLLEVAED 240
DB 195 KAEVYKQP-----SYLVKDFDEKKAITYGELIQDALLSWKTAASPVLGELG-- 242
QY 241 PQOVSTVHLQGLPLVFTVSOQATYEADRTAE--WVAMRLGKAGVLLLRDSLEVNIP 298
DB 243 PETYSG-YITAGIPLAYIFA-ETKEERQGFEEFKFIKHKGSINIYTI----- 290
QY 299 QDANVVKRAEGVPEFLVLDVLLISHVENNMHIE-----EIQEDSDN----- 344
DB 291 -----DAKLYGAHA-GNINLDPSPFAFALIDPPKNAKYPYDQ 327
QY 345 --DMEGPD-----DVODDEVAETVFRD--RKRLPLLELTVELTEFNATVNASDSIV 394
DB 328 SKEVKADIGKFIQDVLDKVEPSIKSEAIPEIQEGPTVVV--AAHYKDLVLDNEKD 384
QY 395 L--FYAGQAVSMALQSYIDVAVKLGKSTMLLTRINCADMSVCTKQNTTEPPIIMKY 452
DB 385 LLEFYAPWCGHCKALAPKYEELASLYKDIPVTLAKID-ATANDV--PDSITGFPITLIF 441
QY 453 KKG--ENPVSTAGMLGKDKLKFQILNRISYPAVITSIQEAEFLSGE 498
DB 442 AAGAKDSPVEYEGSRTEVDLANFVKENG-KHKVDALVDPKKEOESGD 488

RESULT 7
US-09-262-666-4
Sequence 4, Application US/09262666
Patent No. 6346244
GENERAL INFORMATION:

APPLICANT: Hjort, Carsten Mølland
TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 63462440 No. 6346244disk of No. 6346244th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/262,666
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/557,122
FILING DATE: 11-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3980.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 511 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-262-666-4

Query Match 4.2%; Score 172.5; DB 4; Length 511;

Best Local Similarity 19.9%; Pred. No. 2.7e-07; Matches 105; Conservative 107; Mismatches 185; Indels 131; Gaps 28;

QY 15 VMICFYMPVNSLPESLPQKFFSTLQGLNELNVAVPLDYGISAVKVCKEE-ISR 73
DB 48 LVLAFFAPWCGHCKALAP-KY-----EQAAELKEKNIPLVKVCTEEALCR 95
QY 74 YCGEKDLMKAY----LFKGNILREPTDTLFDVNAIVAHVLAFLSEKYYIT--NLE 127
DB 96 DQGEV-----GYPLKIFRGDAVK--PYGARGTEAIVSVWKOST--PAVSPTPENLE 147
QY 128 DLQIENNAKGNIIISYVAIGIPEHRAVMEAGFYGT-----TYQVLTTEIALL 180
DB 148 EIKTMDTI-----VIGIYASDD-----QTANDITTTAESORDNYLEAATSDASIA 194
QY 181 ESIGSEDEVANHLFFHCKLVLDLTOCCRTLMEOPLTTNLINHLFIKTMKAPLLEVAED 240
DB 195 KAEVYKQP-----SYLVKDFDEKKAITYGELIQDALLSWKTAASPVLGELG-- 242
QY 241 PQOVSTVHLQGLPLVFTVSOQATYEADRTAE--WVAMRLGKAGVLLLRDSLEVNIP 298
DB 243 PETYSG-YITAGIPLAYIFA-ETKEERQGFEEFKFIKHKGSINIYTI----- 290
QY 299 QDANVVKRAEGVPEFLVLDVLLISHVENNMHIE-----EIQEDSDN----- 344
DB 291 -----DAKLYGAHA-GNINLDPSPFAFALIDPPKNAKYPYDQ 327
QY 345 --DMEGPD-----DVODDEVAETVFRD--RKRLPLLELTVELTEFNATVNASDSIV 394
DB 328 SKEVKADIGKFIQDVLDKVEPSIKSEAIPEIQEGPTVVV--AAHYKDLVLDNEKD 384
QY 395 L--FYAGQAVSMALQSYIDVAVKLGKSTMLLTRINCADMSVCTKQNTTEPPIIMKY 452
DB 385 LLEFYAPWCGHCKALAPKYEELASLYKDIPVTLAKID-ATANDV--PDSITGFPITLIF 441

QY 453 KKG-ENPVSYAGMLGTDKDLKFIQNLNRSYPVNITSIOEAEYLSGE 498
DB 442 AAGAKDSPYEYEGSRIVEDLANFVKENG-KHKVDALEVPKKEQESOD 488

RESULT 8

US-08-557-122A-3
; Sequence 3, Application US/08557122A
; Patent No. 5879664
; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Malland
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 58796640 No. 5879664disk of No. 5879664th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,122A
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3980.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-557-122A-3

Query Match 4.2%; Score 172.5; DB 2; Length 515;
Best Local Similarity 19.9%; Pred. No. 2,8e-07;
Matches 105; Conservative 107; Mismatches 185; Indels 131; Gaps 28;

QY 15 VIMCIFYMPTVNSLPESLQKYSTLQPGLEELNEAVRPLQDYGISVAKVNCVKEE-ISR 73
DB 48 LVLAEEFAWCHCKALAP-KY-----EQATLEKKNIPLVKVDCTEEBALCR 95
QY 74 YGGEKEDLMKAY----LFGNLLREFPTDTLFDVNAIYAHVLFALLSEVKYIT--NLE 127
DB 96 DQGE-----GYPLKIFFGDLAVK--PYOGARQTEAIVSYWKSL-PAVSPVTPEINLE 147
QY 128 DLQNIENALKGANIIFFSVRAIGIPENRAVMEAGFVYGT-----TYQFVLTEIAL 180
DB 148 EIKTMDKI-----VIGIYASDD-----QTANDIFTFEASQRDNYLFAATSDASIA 194
QY 181 ESIGSEDEVYAHLYFFHCKLVLDLTQOCRRITLMEQPLTTLNHLFIKTKAKAPLLEVAED 240
DB 195 KAEGVKQP-----SIVLYKDFDEKKATYDGEIEDALLSWKTAFTPLVGEIG-- 242
QY 241 POOVSTVHLQGLPLVEIYSQATYEAADRRTAE--VWAMRLKAGVLLLRDSLEVNIP 298
DB 243 PETYSG-YTTAGIPLAYIFA-EIKEREQPTTEEFKIAEKHKSINIYIT----- 290
QY 299 QDANVYFKAEGEVPEVFLVDHVDLIISHVENNHIE-----EIQDEEDN----- 344
DB 291 -----DAKLYGAHA-GNLLDPSKPPAFALIQPEKNAKKYPQD 327
QY 345 --DMGSPDI-----DVQDEVAETVFRD--RRKKLPLELJVELJEEFNATVMSDSIV 394

DB 328 SKEVKAADIGKFIQVDLDDKVEPSIKSEAIPEQOEGPVTYVW--ASHYKDLVDNEKDV 384
QY 395 L--FYAGMQAVSAFLQSYIDVAVKLGKSTMLTIRINCADMSDYCTKQNVTEFPITIKY 452
DB 385 LLEFYAPWCHCKALAPKKEELASLYKDIPETVIKID-ATANDV--PDSITGFPIKLE 441
QY 453 KKG-ENPVSYAGMLGTDKDLKFIQNLNRSYPVNITSIOEAEYLSGE 498
DB 442 AAGAKDSPYEYEGSRIVEDLANFVKENG-KHKVDALEVPKKEQESOD 488

RESULT 9

US-08-557-122A-34
; Sequence 34, Application US/08557122A
; Patent No. 5879664
; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Malland
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 58796640 No. 5879664disk of No. 5879664th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,122A
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3980.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-557-122A-34

Query Match 4.2%; Score 172.5; DB 2; Length 515;
Best Local Similarity 19.9%; Pred. No. 2,8e-07;
Matches 105; Conservative 107; Mismatches 185; Indels 131; Gaps 28;

QY 15 VIMCIFYMPTVNSLPESLQKYSTLQPGLEELNEAVRPLQDYGISVAKVNCVKEE-ISR 73
DB 48 LVLAEEFAWCHCKALAP-KY-----EQATLEKKNIPLVKVDCTEEBALCR 95
QY 74 YGGEKEDLMKAY----LFGNLLREFPTDTLFDVNAIYAHVLFALLSEVKYIT--NLE 127
DB 96 DQGE-----GYPLKIFFGDLAVK--PYOGARQTEAIVSYWKSL-PAVSPVTPEINLE 147
QY 128 DLQNIENALKGANIIFFSVRAIGIPENRAVMEAGFVYGT-----TYQFVLTEIAL 180
DB 148 EIKTMDKI-----VIGIYASDD-----QTANDIFTFEASQRDNYLFAATSDASIA 194
QY 181 ESIGSEDEVYAHLYFFHCKLVLDLTQOCRRITLMEQPLTTLNHLFIKTKAKAPLLEVAED 240
DB 195 KAEGVKQP-----SIVLYKDFDEKKATYDGEIEDALLSWKTAFTPLVGEIG-- 242
QY 241 POOVSTVHLQGLPLVEIYSQATYEAADRRTAE--VWAMRLKAGVLLLRDSLEVNIP 298

```

Db 243 PETSG-YITAGIPLATIFA-ETKEERQFTEEFKFIKHKGINITYI----- 290
Qy 299 QDANVFRAEAGVVEELVLHDVLIISHVENNHIE-----EIOEDEN----- 344
Db 291 -----DAKLYGAHA-GNINLDPKFPAPAIODPENNAKPYDQ 327
Qy 345 --DMEGPD-----DVODEVAETVFRD--RRKRLPLELTVELTEFPNATVMSDSIV 394
Db 328 SKEVAKDIGKFIQDVLDDKVEPSIKSEAIPETOGPTVVV--AHGYKDLVLNEDV 384
Qy 395 L-FYAGQAVSMAPLOSIVAVKLGSTMLTRINCADMSDYCTQNTPEPTIMY 452
Db 385 LLEFYAPMGCHKALAPRYEELASLYKDIPETIAKID-ATANDV--PDSITGFTIKLF 441
Qy 453 KKG--ENPVSYAGMLGTDLKFIQNLNRSYPVNTSIOEAEVYSGE 498
Db 442 AAGAKDSVEYEGSRTEVDLANFVKENG-KHKVDALVDPKKEDESGD 488

```

RESULT 10

```

US-09-262-666-3
; Sequence 3, Application US/09262666
; Patent No. 6346244
; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Malland
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 63462440 No. 6346244disk of No. 6346244th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/262,666
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/557,122
; FILING DATE: 11-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambdiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3980,204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-262-666-3

```

Query Match 4.28; Score 172.5; DB 4; Length 515;

Best Local Similarity 19.98; Pred. No. 2.8e-07; Mismatches 185; Indels 131; Gaps 28;

```

Qy 15 VIMCFYMPYVNSLPELSPQKFFSTLQGLLELNEAVRPLDYGISVAKVCVKEE-ISR 73
Db 48 LVLAFFAPMGCHKALAP-KY-----EQAAELKEKNPIPKVDCTEEEALCR 95
Qy 74 YCGEKDLMAKY---LFGKNIILREFPTDPLFDVNAIVAVLALFSEKKYIT--NLE 127
Db 96 DQGE-----GYPLKIFRGIDAVK--PYOGARQTEAIVSYVXOSL-PAVSPTVPENLE 147

```

```

Qy 128 DLQNIENALGKANITFSYVRAIGIPENHVAWEAGFYGT-----TYQFVLTTEALL 180
Db 148 EKTMDKI-----VIGIYASDD-----QTANDIFTFEASORDNVLFAATSDASIA 194
Qy 191 ESIGSEDEYVAHLFFHCKLVLDLITQOCHRTIMEOPPLTTLNHLFIKTKAPLLEVAED 240
Db 195 KAEGVKQF-----SIVLYKDPDEKKAAYDDEIODALLSWVKTASPTVLGELG-- 242
Qy 241 PQQVSTVLQGLPLVFIVSQATYEADRTAE--WVAMRLLGKRGVILLNDSLEVINP 298
Db 243 PETSG-YITAGIPLATIFA-ETKEERQFTEEFKFIKHKGINITYI----- 290
Qy 299 QDANVFRAEAGVVEELVLHDVLIISHVENNHIE-----EIOEDEN----- 344
Db 291 -----DAKLYGAHA-GNINLDPKFPAPAIODPENNAKPYDQ 327
Qy 345 --DMEGPD-----DVODEVAETVFRD--RRKRLPLELTVELTEFPNATVMSDSIV 394
Db 328 SKEVAKDIGKFIQDVLDDKVEPSIKSEAIPETOGPTVVV--AHGYKDLVLNEDV 384
Qy 395 L-FYAGQAVSMAPLOSIVAVKLGSTMLTRINCADMSDYCTQNTPEPTIMY 452
Db 385 LLEFYAPMGCHKALAPRYEELASLYKDIPETIAKID-ATANDV--PDSITGFTIKLF 441
Qy 453 KKG--ENPVSYAGMLGTDLKFIQNLNRSYPVNTSIOEAEVYSGE 498
Db 442 AAGAKDSVEYEGSRTEVDLANFVKENG-KHKVDALVDPKKEDESGD 488

```

RESULT 11

```

US-09-262-666-34
; Sequence 34, Application US/09262666
; Patent No. 6346244
; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Malland
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 63462440 No. 6346244disk of No. 6346244th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/262,666
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/557,122
; FILING DATE: 11-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambdiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3980,204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 515 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-262-666-34

```

```
Query Match          4.2% Score 172.5; DB 4; Length 515;  
Best Local Similarity 19.9%; Pred No. 2,8e-07;  
Matches 105; Conservative 107; Mismatches 185; Indels 131; Gaps  
  
OY      15 VIMCIFFPTVNSLDELSPQKYFSTLOQLLENEAVRPLADYGISVANVCKEE--ISR 73  
        :|::||::||::||:|::||:|::||:|::||:|::||:  
Db       48 LVLAFFPAPWCHCKALAP-KY-----EQATELKKNIPLVAKDCDEEALCR 95  
  
OY      74 YCGRKDLMAK----LFGRNLLREPFDTLPDVNAIYAHLFALLESEVKYT--NLE 127  
        :|::||::||:|::||:|::||:|::||:|::||:|::||:  
Db       96 DGVG-----GPULKIFRGDAVR--PYGAROTEAIVSYWKSL-PAYSVPENLE 147  
  
OY     128 DLONTENALKGANIIIFSFRVRAIGIEPRHVAWEAGFYGT-----TYOFVLTEIAL 180  
Db    148 EKTMDKI-----VGITIASD-----QTANDITFPASORDNLFAATSASTA 194  
  
OY     181 ESIGSEDVEYAHLYEFHCXLDLTQQCRRTLMEOPLTTLINILFIKTKARPLLFEVD 240  
        :|::||::||:|::||:|::||:|::||:|::||:|::||:  
Db    195 KAEVKKQP-----SIVLKPDFDEKATYYDGIEEDALLSMWKTAFASTPLSGT-- 242  
  
OY     241 PQOVSTVHQLGLPLVFYSQAATEADRRTAE--VWAMRLLGKACVLLLRDSLEVNI 298  
        :|::||::||:|::||:|::||:|::||:|::||:|::||:  
Db    243 PETSYG-YITAIPLAYIFA-EFKEREQEFTPEEKIAKHKSINIYVI----- 290  
  
OY     299 QDANVFRAEGEVVEFLHDVOLIISHENNHI-----EIODEDN----- 344  
Db    291 -----DAIKYGAHA-GNLMLDSPKRPAFIQDPKNAKPIYO 327  
  
OY     345 --DMGGPDI---DVODEVAETVFRD--RRKLPLELTVELTEEPENATVMASDISV 394  
        :|::||::||:|::||:|::||:|::||:|::||:|::||:  
Db    348 SKEVKAADIGKFIDVLDLKDPEPSIKEAIPEIQEGPVVVV---ASHXYKDVLDERKV 384  
  
OY     395 L-FVAGNQVSMARLOSVIDVAVYLKGISTMULIRINADMDSVCTKKNTEFFLIKKY 452  
        :|::||::||:~::~||:~::~||:~::~||:~::~||:~::~||:  
Db    385 LEEFYAPMGCHKALAPKYEELASLYKDIPEVTIAKD-ATANDV--PDSTGFPPKLF 441  
  
OY     453 KKG--ENEVSYAGMLGTKDKLDKFIOLNRISYPVNITSIOEAELYSGE 498  
        |::||::||:~::~||:~::~||:~::~||:~::~||:~::~||:  
Db    442 AAGAARDPEYEGSRIVEDLANFVENKENG-KHKVDALEYPKRQESGD 488  
  
RESULT 12  
US-08-627-907A-2  
Sequence 2, Application US/08627907A  
Patent No. 6060302  
GENERAL INFORMATION:  
APPLICANT: HIRANO, Naoto  
APPlicant: HIRAI, Hisamaru  
TITLE OF INVENTION: HUMAN PHOSPHOLIPASE C- $\alpha$ -ALPHA AND DNA  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESSES:  
ADDressee: c/o FISH & NEAVE  
STREET: 1251 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10020  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/M5-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/627,907A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 5-238402  
FILING DATE: 24-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP94/01572  
FILING DATE: 22-SEP-1994
```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:13:27 ; Search time 16 Seconds
(Without alignments)
2324.769 Million cell updates/sec

Title: US-09-847-046-2
Perfect score: 4142
Sequence: 1 MSGGFNFRVIGISFVIMCIF.....FKEAKSFRDKEIGSKSVN 807

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PC7_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description |
|------------|-------|-------------|-----------|-----------------------|--------------------|
| 1 | 4142 | 100.0 | 807 | 10 US-09-847-046-2 | Sequence 2, Appl |
| 2 | 177 | 4.3 | 496 | 9 US-10-108-605-51 | Sequence 51, Appl |
| 3 | 169.5 | 4.1 | 515 | 10 US-09-759-010-8 | Sequence 8, Appl |
| 4 | 169.5 | 4.1 | 544 | 10 US-09-925-301-869 | Sequence 869, App |
| 5 | 132 | 3.2 | 468 | 10 US-09-925-300-1661 | Sequence 1661, App |
| 6 | 132 | 3.2 | 769 | 10 US-09-788-657-15 | Sequence 15, Appl |
| 7 | 129 | 3.1 | 396 | 9 US-09-854-133-436 | Sequence 436, App |
| 8 | 129 | 3.1 | 432 | 10 US-09-738-973-436 | Sequence 436, App |
| 9 | 128 | 3.1 | 432 | 9 US-09-978-295A-90 | Sequence 90, Appl |
| 10 | 128 | 3.1 | 432 | 9 US-09-978-697-90 | Sequence 90, Appl |
| 11 | 128 | 3.1 | 432 | 9 US-09-978-192A-90 | Sequence 90, Appl |
| 12 | 128 | 3.1 | 432 | 9 US-09-999-832A-90 | Sequence 90, Appl |
| 13 | 128 | 3.1 | 432 | 9 US-09-978-189-90 | Sequence 90, Appl |
| 14 | 128 | 3.1 | 432 | 9 US-10-174-590-74 | Sequence 74, Appl |
| 15 | 128 | 3.1 | 432 | 9 US-10-176-758-74 | Sequence 74, Appl |
| 16 | 128 | 3.1 | 432 | 9 US-10-175-737-74 | Sequence 74, Appl |
| 17 | 128 | 3.1 | 432 | 9 US-10-173-706-74 | Sequence 74, Appl |
| 18 | 128 | 3.1 | 432 | 9 US-10-175-738-74 | Sequence 74, Appl |
| 19 | 128 | 3.1 | 432 | 9 US-10-175-752-74 | Sequence 74, Appl |

| | | | | | | |
|----|-----|-----|-----|---|------------------|-------------------|
| 20 | 128 | 3.1 | 432 | 9 | US-10-176-482-74 | Sequence 74, Appl |
| 21 | 128 | 3.1 | 432 | 9 | US-10-176-757-74 | Sequence 74, Appl |
| 22 | 128 | 3.1 | 432 | 9 | US-10-176-913-74 | Sequence 74, Appl |
| 23 | 128 | 3.1 | 432 | 9 | US-10-180-552-74 | Sequence 74, Appl |
| 24 | 128 | 3.1 | 432 | 9 | US-10-180-557-74 | Sequence 74, Appl |
| 25 | 128 | 3.1 | 432 | 9 | US-10-173-700-74 | Sequence 74, Appl |
| 26 | 128 | 3.1 | 432 | 9 | US-10-174-572-74 | Sequence 74, Appl |
| 27 | 128 | 3.1 | 432 | 9 | US-10-174-579-74 | Sequence 74, Appl |
| 28 | 128 | 3.1 | 432 | 9 | US-10-174-587-74 | Sequence 74, Appl |
| 29 | 128 | 3.1 | 432 | 9 | US-10-174-588-74 | Sequence 74, Appl |
| 30 | 128 | 3.1 | 432 | 9 | US-10-175-739-74 | Sequence 74, Appl |
| 31 | 128 | 3.1 | 432 | 9 | US-10-175-740-74 | Sequence 74, Appl |
| 32 | 128 | 3.1 | 432 | 9 | US-10-175-743-74 | Sequence 74, Appl |
| 33 | 128 | 3.1 | 432 | 9 | US-10-176-488-74 | Sequence 74, Appl |
| 34 | 128 | 3.1 | 432 | 9 | US-10-176-492-74 | Sequence 74, Appl |
| 35 | 128 | 3.1 | 432 | 9 | US-10-176-747-74 | Sequence 74, Appl |
| 36 | 128 | 3.1 | 432 | 9 | US-10-176-750-74 | Sequence 74, Appl |
| 37 | 128 | 3.1 | 432 | 9 | US-10-176-985-74 | Sequence 74, Appl |
| 38 | 128 | 3.1 | 432 | 9 | US-10-176-987-74 | Sequence 74, Appl |
| 39 | 128 | 3.1 | 432 | 9 | US-10-176-991-74 | Sequence 74, Appl |
| 40 | 128 | 3.1 | 432 | 9 | US-10-176-997-74 | Sequence 74, Appl |
| 41 | 128 | 3.1 | 432 | 9 | US-10-176-999-74 | Sequence 74, Appl |
| 42 | 128 | 3.1 | 432 | 9 | US-10-184-658-74 | Sequence 74, Appl |
| 43 | 128 | 3.1 | 432 | 9 | US-10-173-695-74 | Sequence 74, Appl |
| 44 | 128 | 3.1 | 432 | 9 | US-10-173-697-74 | Sequence 74, Appl |
| 45 | 128 | 3.1 | 432 | 9 | US-10-173-705-74 | Sequence 74, Appl |

ALIGNMENTS

```
RESULT 1
US-09-847-046-2
Sequence 2, Application US/09847046
Patent No. US20020068036A1
GENERAL INFORMATION:
APPLICANT: Hevezi, Peter
APPLICANT: Mack, David
APPLICANT: Gish, Kurt C.
APPLICANT: Wilson, Keith E.
TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF PROSTATE AND/OR BREAST CANCER,
TITLE OF INVENTION: COMPOSITIONS, AND METHODS OF SCREENING FOR PROSTATE CANCER, AN
FILE REFERENCE: A-69199-1/DJB/JJD/AMS
CURRENT FILING DATE: 2001-04-30
PRIOR APPLICATION NUMBER: US/09/847,046
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: US 09/733,288
PRIOR FILING DATE: 2000-10-13
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 807
TYPE: PRT
ORGANISM: Homo sapiens
US-09-847-046-2
Query Match 100.0%; Score 4142; DB 10; Length 807;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSGGFNFRVIGISFVIMCIFPTVNSLPELSPOKFFSTLQGLLELNEAVRPDDYGIS 60
DB 1 MSGGFNFRVIGISFVIMCIFPTVNSLPELSPOKFFSTLQGLLELNEAVRPDDYGIS 60
QY 61 VAKVVCVEEISRYGCKEKDKAVLEFGNILLREPTDTLFDVAIVAHVIFALLFSEV 120
DB 61 VAKVVCVEEISRYGCKEKDKAVLEFGNILLREPTDTLFDVAIVAHVIFALLFSEV 120
QY 121 KYTNLEDLQNTENALKKRANIIFSVYRAIGIPEHRAVMEAGFYVGTYYQFVLTTEIAL 180
DB 121 KYTNLEDLQNTENALKKRANIIFSVYRAIGIPEHRAVMEAGFYVGTYYQFVLTTEIAL 180
```

| | | | |
|----|-----|--|-----|
| Qy | 181 | ESISEDEVEYAHLEFHCKLVDLTQOCRRJLMOPLETTINHLFIETKMAPLLEVAED | 240 |
| Db | 181 | ESISEDEVEYAHLEFHCKLVDLTQOCRRJLMOPLETTINHLFIETKMAPLLEVAED | 240 |
| Qy | 241 | POQVSTVHLQJGLPVLVFSQOATYEADRRTAEWVAANRLGKAGVLLLLDSLEVNIPQD | 300 |
| Db | 241 | POQVSTVHLQJGLPVLVFSQOATYEADRRTAEWVAANRLGKAGVLLLLDSLEVNIPQD | 300 |
| Qy | 301 | ANVYFKRAEEGVPAPEFLVYLBHVDLIISHVENNMHIEEIOEDENDMGPDIDVQDDEVAE | 360 |
| Db | 301 | ANVYFKRAEEGVPAPEFLVYLBHVDLIISHVENNMHIEEIOEDENDMGPDIDVQDDEVAE | 360 |
| Qy | 361 | TVFDRKRRKLLELELVETEETFAATYMASSIVLYFAGMOAVSMALQSYIDVAYKLG | 420 |
| Db | 361 | TVFDRKRRKLLELELVETEETFAATYMASSIVLYFAGMOAVSMALQSYIDVAYKLG | 420 |
| Qy | 421 | TSTMLLTRINCADMSVCTKQNTVEPPIKMYKKGKNPVSAGMLGTRDKLLEFIQJLRIS | 480 |
| Db | 421 | TSTMLLTRINCADMSVCTKQNTVEPPIKMYKKGKNPVSAGMLGTRDKLLEFIQJLRIS | 480 |
| Qy | 481 | YPVNTSIOEAEVYLSGELYKDLIYSSVSYLGIFSPMTAKDESEAGNYLKGVTYG | 540 |
| Db | 481 | YPVNTSIOEAEVYLSGELYKDLIYSSVSYLGIFSPMTAKDESEAGNYLKGVTYG | 540 |
| Qy | 541 | IYSEEDVLLSTKYAASHPALLARHTEGKIYESIPLASTHANOQIVQIITDALLEMPEIT | 600 |
| Db | 541 | IYSEEDVLLSTKYAASHPALLARHTEGKIYESIPLASTHANOQIVQIITDALLEMPEIT | 600 |
| Qy | 601 | VENLPSTYRLOKPLLIIFSQSTQVNPQYKAKALLTVYKQYIDLSFPCMLNKNTPVNGIL | 660 |
| Db | 601 | VENLPSTYRLOKPLLIIFSQSTQVNPQYKAKALLTVYKQYIDLSFPCMLNKNTPVNGIL | 660 |
| Qy | 661 | RAYEDPLPEPLVLVYNHSGQYFAFPSPDAIIEENMLVLMKLEJGLENNHTIIPAGE | 720 |
| Db | 661 | RAYEDPLPEPLVLVYNHSGQYFAFPSPDAIIEENMLVLMKLEJGLENNHTIIPAGE | 720 |
| Qy | 721 | WKPPPLPAYDLMSIDAATSQGRTRKRVPCMKETVDQENDKEQHEDEKSAVAREPIETLRK | 780 |
| Db | 721 | WKPPPLPAYDLMSIDAATSQGRTRKRVPCMKETVDQENDKEQHEDEKSAVAREPIETLRK | 780 |
| Qy | 781 | HMNSNMFKAEKSPRRDKELGCSKVN | 807 |
| Db | 781 | HMNSNMFKAEKSPRRDKELGCSKVN | 807 |

```

RESULT 2
US-10-108-605-51
: Sequence 51, Application US/10108605
: Patent No. US20020160934A1
: GENERAL INFORMATION:
: APPLICANT: Broadus, Julie
: APPLICANT: Stam, Lynn
: APPLICANT: Bachmann, Jane
: APPLICANT: Kamdar, Kim
: TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODED
: TITLE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF
: FILE REFERENCE: 31133B
: CURRENT APPLICATION NUMBER: US/10/108,605
: CURRENT FILING DATE: 2002-03-27
: PRIOR APPLICATION NUMBER: US 09/761,142
: PRIOR FILING DATE: 2001-01-16
: PRIOR APPLICATION NUMBER: US 60/176,418
: PRIOR FILING DATE: 2000-01-14
: NUMBER OF SEQ ID NOS: 361
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 51
: LENGTH: 496
: TYPE: PRT
: ORGANISM: Drosophila melanogaster
US-10-108-605-51

```

| | Best Local Similarity | 19.9%; | Pred. No. 1.6e-06; | |
|---------|-----------------------|--|---------------------------------------|--------------------------------------|
| Matches | 99; | Conservative | 96; | Mismatches 180; Indels 122; Gaps 25; |
| QY | 373 | ELTVELTEETFNATV----- | MASDIYL--FYAGQAVSNMFLQSYIDVAVKL-KGIST | 423 |
| Db | 20 | EAEVVEEGLVATVADNFQKLIADNEFVLEVEYAPWCCHKALAEYAKAAQQLAEKSP | | 79 |
| QY | 424 | MLTPTINADMSDVCOTKQVTEFPPIIKMKKGKGNPVSAVGMJGTGDLKFIQLNRIISYV | | 483 |
| Db | 80 | ILKAVVDAITESELAEQYAVRGYPTPLKFFRSG-SFEYESSGGQADDIIMVTKTGPPAK | | 138 |
| QY | 484 | NITSIQEAEVYLSGELYKQDLILYSVSVLGIESPTMKTAKEDEFSAGNYLKYVITGIYS | | 543 |
| Db | 139 | DLTSVADAEOFL-----KD-----NEIATIGFEKDLSEEAQKFTFYVANAIDSEFV-GVSS | | 188 |
| QY | 544 | EDVDVLLSKRYKVAASIPALLLAHTTGKLTISTPLASTHADIYQIITDALLEFPEITVGN | | 603 |
| Db | 189 | NADVI--AKYFAKONGVYLFKPFEDK-----KSFE--GELNEBN | | 224 |
| QY | 604 | LPSTYRFL-PLLIIFSDDGTVPQY---KKAILTVKQKYLDSETPCWLNLKNTPVGKG | | 658 |
| Db | 225 | LKKFAQVQSLPIVDNFNHSASKIRGSGIKSHLFFVSNE-----CG | | 266 |
| QY | 659 | ILRAVFDPLPLPL-----LVLVNLHSGQ---VFAF----- | | 687 |
| Db | 267 | HLKXVDPFLKELAKKYYRDDILFVTISSDEEDTRIFEEFGMKKEEVPTRILKLEEDMAK | | 326 |
| QY | 688 | --PSQAIIIEEMLVYLMLKK-LEAGLENNHTI--LPAQEW-KPP--LPAYDFLSM-IDAA | | 737 |
| Db | 327 | YKPEEDDLASLIEIAFLKFKELDGKLIKHLISGLP-EDMDKRPVAVIVSSNESVALKGS | | 385 |
| QY | 738 | TSQRGTRKVP---KCMKETDVOENDKEQHEDEKSAV-----RKPEIETLRKHHNRSSW | | 787 |
| Db | 386 | KSVLVEFYAPWCGHOKOLAPIYDOLAEEKYKQDNEDIVIAKMDSMTANELESIKISSPTIKY | | 445 |
| QY | 788 | FKEAEK---SFRDKEL 801 | | |
| Db | 446 | FKKEDNKVYDFNLDRTL 462 | | |

```

RESULT 3
US-09-759-010-8
: Sequence 8, Application US/09759010
: Patent No. US20010034042A1
: GENERAL INFORMATION:
: APPLICANT: Sivantova, Pramod K.
: TITLE OF INVENTION: COMPLEXES OF PEPTIDE BINDING FRAGMENTS OF HEAT-SHOCK
: TITLE OF INVENTION: PROTEINS AND THEIR USE AS IMMUNOTHERAPEUTIC AGENTS
: FILE REFERENCE: 8449-135
: CURRENT APPLICATION NUMBER: US/09/759,010
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 8
: LENGTH: 515
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-759-010-8

Query Match          4.1%; Score 169.5; DB 10; Length 515;
Best Local Similarity 20.6%; Pred. NO. 7e-06;
Matches 111; Conservative 89; Mismatches 193; Indels 145; Gaps 27;

OY 15 VMCLFVMPYPTVNSLDLSPOKYEFTSLQGLLELNVAVRPLDQYGISVAKNVCREEISRY 74
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
DB 44 IYLAIEFPAWCGCHKALAP-KT-----EEAATELKKNIPLVAKVDTAEE---- 87
    :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

OY 75 GSEKEDLMKAYLEFGKNILREF-----PTDLEFVAIVAHVLFAL--FSEVKYITN 125
    |:::  |:::  |:::  |:::  |:::  |:::  |:::  |:::  |:::  |:::
DB 88 -----DLCRSQGEVEYPTLTKIFRGVDSSKPYGARGQTSIESYMLKQSLPAYSSVNE-EN 141
    |:::  |:::  |:::  |:::  |:::  |:::  |:::  |:::  |:::  |:::

OY 126 LEDDIONENALKGKNNIIFSYVAIRGIPBHRVAMEAGFYGTTFYVLTLETALLSESTGS 185
    |:::  |:::  |:::  |:::  |:::  |:::  |:::  |:::  |:::  |:::

```

Db 142 LEEIKTMDKI-----VVGIIYIPSDDEFTQAFEKYEASORDNYLFAATDAALAKSEGV 195
QY 186 EDVEYAHLY--FFHCKIYLD--LFOQCRRILMEQPLTTLNHLFIKTKMAKPLTVEADP 241
Db 196 EOPSTIV-LYKDFDEKKANYDGEIEDEA-----IHSWKASSTPLVGEIG--P 239
QY 242 QOVSTVHLQGLPLVEFIYSQOATYEADRRTAEM-----VAMRLIG-KAGV 285
Db 240 EYASG-YIGAGVPLAYIFA-ETKEEREKYTEDFKIQAOKHKAINIATIDAKMFGAHAGN 297
QY 286 LLLRLDSLEVINIPD-----ANVYKRAEE--GVPEFLVLDVDLIISHVENNHIEIIO 339
Db 298 LNLDSQKPPAPALIDPPAKNARYPPQAKELNADEVEKEFI---QDVLDSKVEPSIKSEPP 354
QY 340 EDEDNDMEGPDIDVQDDDEVAETVFRDRKRKLPLELTVELTEETEFATYMASDIYL--FY 397
Db 355 ESQ-----EGP-----VTV--VVAHSYDLYINDKDYLLFEY 385
QY 398 AGWQAVMAFLQSYIDVAVKLGSTMLTRINCADMSDVCTK-----QNV 443
Db 386 APWGHCKALAPKYDELAALY-----ADHPDLAKVITIAKIDATANDVPPI 432
QY 444 TEFPILKRYKKG--ENPYSYAGMGTGKDLKFLIOLNRISTYPNITISIOAEYLSGEL 499
Db 433 TGFPTLRLYPAGAKDSPIEYSGSRTVEDLANFVKENG-KHNYDALNVASEETOEGGDV 489

RESULT 4

US-09-925-301-869

Sequence 869, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA106
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05882
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 869
LENGTH: 544
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (7)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (9)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-869

Query Match

4.1%; Score 169.5; DB 10; Length 544;

Best Local Similarity 22.6%; Pred. No. 7.6e-06;
Matches 64; Conservative 59; Mismatches 127; Indels 33; Gaps 10;

QY 376 VELTEEFNATVMAASD-----IVLFYAGWQAVMAFLQSYIDVAVKLGSTMLTRINC 431
Db 67 LELDDNENESISDTGASGLMLVEFFAPWGHCKRLAPEYEAARLKG--IVPLAKYDC 124
QY 432 ADMSDVCTKQNVTEPPIIKMKKGENPVSYAGMLGKDLKFLIOLNRISTYPNITISIOEA 491
Db 125 TANNTCKKXYGVSYPILKIRDEEAGAYDGPRTADIVS--HLKQAGPASYDLPTREE 182
QY 492 EYLSGELKDLILYSSVYGLFSPMTKAKDEPSEAGNLKGVITIGIYSEEBYLLS 551
Db 183 E-----FKFISDKDASIVGFDDSEAHSEFLKASNLRD--NYRAHNTVESLV 232
QY 552 TKVAAAPLALLAR--HTEGKIESIPLASTHAQ-----DIVOIITDALLEMPPEITVENLP 605

Db 223 NEYDNGEEIILFRSHLTKNFEDRTVAYTEQKMTSGKIKKFIQENIFICPHMETDNKD 292
QY 606 SYFRQ-KPLLLFSD-----GTVPQYKALLTLVKQYLD 642
Db 293 ---LIQKDLIAYVDYDEKNAKGSNWRNRNVMYAKKFLDA 332

RESULT 5

US-09-925-300-1661

Sequence 1661, Application US/09925300
Patent No. US20020151681A1
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Steve Ruben,
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA101
CURRENT APPLICATION NUMBER: US/09/925,300
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1890
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1661
LENGTH: 468
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (2)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1661

Query Match

3.2%; Score 132; DB 10; Length 468;

Best Local Similarity 22.5%; Pred. No. 0.0079;
Matches 112; Conservative 78; Mismatches 211; Indels 96; Gaps 25;

QY 15 VIMCIFYMPTVMSLPBELSQKYSFSLQGLLELNKAVRPLQDYGISAVKNCVKEISY 74
Db 19 IILVEFYAPWGHCKRLAPE-----YEKAKELSRSP-----IPLAKVAT----- 61
QY 75 CGKEKDLKAYLFKGNILNREPTDLPDVN-----ATAHVALLLFSEVKYITNLED 128
Db 62 --AETDLAKRFVSGYPILKIRKCRPYDYNCPREKYGIVDYMGQSPSEKILF---- 115
QY 129 LQINENAK-GRANITFSYVRAIGIPENRAVAGFVYGYTOF--VLTTEIALLLESIGS 185
Db 116 LKQVGFLEKDGDDVITIGFKGESDPAYQOQDAAANNLRDEYKFFHHTSTELAKFLKYSQ 175
QY 186 EDVEYAHLYFFHCKIYLDLFOQCRRILMEQPLTTLN--IHLFIKTKMAKPL--TEVAEDP 241
Db 176 GOLVMOPEKFOSK-----YEPRSHMMDVQGSTODSAIKDVLKALYALVGHRRKVSND 229
QY 242 QOVSTVHLQGLPLVEFIYSQOATYEADRRTAEMVAMRLIGKAGVLLRLDSLEVINIPDA 301
Db 230 KRYTR-----RPLV-VVYISVDFSDYRAAQF-WR-----SKYLEVAKDPE----- 270
QY 302 NVFKEAEEGVVEFLVLDVDLIISHVENNHIEIODEDNDMEGPDIDVQDDDEVAET 361
Db 271 -YTFILADE-----EDYAGEVNDGLISEGSEDVNAALIDESGKKFAMEPE-EPDSOTLREF 324
QY 362 VFRDRKRKL-----PLELTVELTEEFNATVMAASD-----IVLFYAGWQAVS 404
Db 335 VTAFKKGKLYKIKVKSQPPYKNNKGVKVVV--GKTFPSIIVDPKKVDLIEFYAPWGHCK 381
QY 405 MAFLQSYIDVAVKLGSTMLTRINCADMSDV-CTKQNVTEPPIIKMKKGEN--NPVS 461
Db 382 KOLEPVIYSLAKKRYGQGLVIYAKMD-ATANDVPDRKYVGEFPIITYAPSGDKKNPVKF 440
QY 462 AGMLGTXD--LLEFIQ 475

Db 441 EG--GDRDLHUSKITE 455

RESULT 6

US-09-788-657-15

Sequence 15, Application US/09788657

Patent No. US20020123149A1

GENERAL INFORMATION:

APPLICANT: Nicolaides, Nicholas

APPLICANT: Sassi, Philip

APPLICANT: Kinzler, Kenneth

APPLICANT: Grasso, Luigi

APPLICANT: Vogelstein, Bert

TITLE OF INVENTION: Methods for generating hypermutable

TITLE OF INVENTION: Yeast

FILE REFERENCE: 01107.00097

CURRENT APPLICATION NUMBER: US/09/788,657

CURRENT FILING DATE: 2001-02-21

PRIOR APPLICATION NUMBER: 60/184,336

PRIOR FILING DATE: 2000-02-23

NUMBER OF SEQ ID NOS: 25

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 15

LENGTH: 769

TYPE: PR

ORGANISM: Saccharomyces cerevisiae

US-09-788-657-15

Query Match 3.2%; Score 132; DB 10; Length 769;

Best Local Similarity 19.2%; Pred. No. 0.017;

Matches 161; Conservative 135; Mismatches 300; Indels 244; Gaps 44;

59 ISVAKVNCVKEISRYCGEKDLMKAYLFGNLLREFPTDLEFVNAIYAVHFLAFPS 118

21 IISPNALKEKEMENSIDANATMIDILVEKGGIKVLI--TDNSGINKADLPIL--CERET 78

119 EVKYNLEDLQNIEN--ALKGRNITFSYVRAI---GPEHNAVNAAGVYVITQ--- 170

79 TSK-LQKFDLSDIOIYQYGGFEGALASISHVARTVTTKVEDCANKRVSAEKGMLSEPK 137

171 -----FVLTEI-----ALLESIGSEDEVEYAH-----YFPHCKLVLDTQOCRR- 210

138 PVAKGDTIIVEDLEFNIPSRRLARSHNDESKILDVGRRAHISK---DIGFCKKF 194

211 -----TLMQPLTTNLNLH---FIKTKAPLTL---EVAED-----POGVSTVHL--OL 251

195 GDSNYSLSYKPSYTYODRIYFVFNKSVASNLITFHISKYEDLLESVDGVCMLNFIISK 254

252 GLPIVATVGO--ATYADARTAEWVAMRLTKA-----GVLLLLRDLSEVNIIPDANV 303

255 SISLIEFINNRVLVCDLRRALNSVSNLPRKGFPPYLGIVY--DPAAY---DVNV 307

304 VKRAEEGVPEFLVLDVLIISHVENNMHIEIOEDEDDMEGPDIDVODE---VA 359

308 HPTKRE---VRL--SQDEIIETKIANOLHAEALDHSRPFKASISTNNPESLIPN 360

360 ETVFDRDRKK-LPELITVELTEBETENATYMAS---DSIVLEVAGQAVSMALOSYIDV 414

361 DTIESRNRKSLRQAVVENSYTTANSQLRKAKROENKLVRIDAS--QAKITSPLSS--SQ 417

415 AKLKGISTM-----LRLINCAHMSDVCT-----KQNT 444

418 QNFESSSTKROLSEPKVNVSHSOAEKRLTNSESPDANTINDNDLKDPKRRQKKG 477

445 EF--PIIKYKKGENVVSAGMLGTDLKFTQLNRSVPVNTSIOEAEVYLSGELYD 502

478 DKVPSIADDEKNALDISDG-----YIRVPERVNVNLTSIKKLRKYDQSIHRE 528

503 LI-LYSSVVLGLFPTMTAKEDSEAGNYLKGIVITGISEEDVLLSTFYAASLPAL 561

529 LNDIFANLVGVVDEERLAA-----IQHDKLLIDYGSVCYEL 569

562 -----LLARHTEGKTESIPLASTHAODIVQIITDALLEMPETIVE----- 602

Db 570 FVOIGLTDFANFGKIN---LQSTNVSQ--DIVLYNLSEFDELNDNDASKKEKITSIMDS 624

603 -NLPSYERLQKPLLEFSDQVNPQYKKAILETVKOKYLDSEFPWMLKPNVGVGILR 661

625 SMUNETYSIE-----LVNGDLND-----LKSVM-----LKSPL---LTK 657

662 AVFDPPLPLPLVNLVHSGGVFAFPSPDAILIENLVMLKKLEAGLENNHITILPAQEM 721

658 GYIPSLVKLPFFI-----YRLGKEVDEDEQECID-----GILREIALLYIPDM 701

722 KPLPLAVDPLSMDATSQRGRTKVKCKMETDVQENQKEDKSAVKEPIETLRIRH 781

702 VPKVDITLD-----ASLSEDEKAQFTN---RKEHISL--LEH 733

RESULT 7

US-09-854-133-436

Sequence 436, Application US/09854133

Publication No. US20020183499A1

GENERAL INFORMATION:

APPLICANT: Lodes, Michael J.

APPLICANT: Mohamath, Raedoh

APPLICANT: Henderson, Robert A.

APPLICANT: Benson, Darin R.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.475C10

CURRENT APPLICATION NUMBER: US/09/854,133

CURRENT FILING DATE: 2001-05-11

NUMBER OF SEQ ID NOS: 735

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 436

LENGTH: 396

TYPE: PR

ORGANISM: Homo sapiens

US-09-854-133-436

Query Match 3.1%; Score 129; DB 9; Length 396;

Best Local Similarity 20.4%; Pred. No. 0.011;

Matches 79; Conservative 54; Mismatches 134; Indels 120; Gaps 15;

379 TEETFNATYMASDSIVLYFAGQAVSMALOSYIDVAVKLG--ISTMLLTRINCAHMSD 436

31 TADMFTHGIOSAHEFVWFAPMCGHCQRLOPTWMDLGDYNSMEDAKYVAVYDCTAHS 90

437 VCTKQVTEFPPIIKYKKEPNVSAGMLGTDLKFTQLNRSVPVNTSIOEAEYLS 496

91 VCSAGVGRVPTILKLPQGEAVKYGPPDFOTLENMMLQTLNEEFV--TPPEVEPPSA 148

497 GELYKDLIYSSVS-----VIGL-FSPMTKTAK 523

149 PELKQGLYELASNFELHVAQGDHFIKFPAPMCGHCKALAPTWEOALGLEHSEYVICK 208

524 EDSEF-----AGNYLKGYITIGISEEDVLLSTKYAASLPALLARHTEGK----- 570

209 VQCTQHYELCSNOVQY-----PILWFR--DGKRVQYKCK 244

571 -----IES--IPLASTHAODIVQIITDALLEMPETIVNLSPYRFLPLILFSD 620

245 RULESIREYVESQLORTETGATEIVPSBAVLAEP-----ADGIVLALTE 293

621 GTVNPQYKKAILETVK-----OKYLDSEFPW--LNLKNTPVGRGILRAYFD----- 665

294 NTFDDITAGT--TFIKFYAPMCGHCKITLAPTWBELSKKEFPOLAGVAKIAEVCTAERNIC 352

666 ---PPLPLPLV-----VNLHSGQ 683

353 SKYSVGRVPTLLPLFGKKVSEHSGR 379

RESULT 8

US-09-738-973-436
; Sequence 436, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Flinn, Steven P.
; APPLICANT: Mohamath, Raedoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indirias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 436
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-973-436

Query Match 3.1%; Score 129; DB 10; Length 396;
Best Local Similarity 20.4%; Pred. No. 0.011;
Matches 79; Conservative 54; Mismatches 134; Indels 120; Gaps 15;

Qy 379 TEEFTNATVMSDSIVLYTAGQAVSMAFLQSYIDVAVKLGK--ISTMLTINCADMSD 436
Db 31 TADMTHGIQSAHFVMEFAPCGHCORLQPTWMDLGKYNSEMAKVVAVKVDCTAHS 90
Qy 437 VCTKNTVEFPILKMKYKGENPVSYAGMLGTDLKFIQNLNRSYVNTSIOEAEYLS 496
Db 91 VCSAGVAGYPTLKLFKFGQEVAVKQGRDPTLNMMLQTLNEBPV--TPREVEPPSA 148
Qy 497 GELYKDLILYSVS-----VIGL-FSPYTKAK 523
Db 149 PELKQGLYELASNFELHVAQGDHFKEFAWCGHCKALAPTWEDALGLEHSEYKICK 208
Qy 524 EDFSE-----AGNYLKGVITIGYSEEDVLLSTYASLPLLLARHTEGR----- 570
Db 209 VDCIOTHYELCSGNQVRGY-----PTLWFR--DGKKVDQYKCK 244
Qy 571 -----IES-IPLASTHADIVQIITDALLEMPETTVENLPSYFLQKPLILFSD 620
Db 245 RDLESRLREVESQLORTETGATEYTPSEAPVLAEP-----ADGTYIALTE 293
Qy 621 GTVNPQYKAILTLVK-----OKYLDSTPCW--LNLKNTPVGRGILRAYFD----- 665
Db 294 NTFDDTIAEGI-TFIKFAWPGCHCKTLAPTWEELESKKEFPOLAGVKIAVEDCTAERNIC 352
Qy 666 ---PLPLPLVL-----VNLHSGG 683
Db 353 SKYSVRGYPTLLFRGKKVSEHSGR 379

RESULT 9
US-09-978-295A-90
; Sequence 90, Application US/09978295A
; Patent No. US20020156006A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Bostein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon

; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Kuo, Sophia S.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James.
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Shelton, David L.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2630PIC11
; CURRENT APPLICATION NUMBER: US/09/978,295A
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 09/918585
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064249
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077641
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078004
; PRIOR FILING DATE: 1998-03-13
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078936
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079920
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/079923

[illegible]

```

1 PRIOR APPLICATION NUMBER: 60/083500
2 PRIOR FILING DATE: 1998-04-29
3 PRIOR APPLICATION NUMBER: 60/083742
4 PRIOR FILING DATE: 1998-04-30
5 PRIOR APPLICATION NUMBER: 60/084366
6 PRIOR FILING DATE: 1998-05-05
7 PRIOR APPLICATION NUMBER: 60/084414
8 PRIOR FILING DATE: 1998-05-06
9 PRIOR APPLICATION NUMBER: 60/084441
10 PRIOR FILING DATE: 1998-05-06
11 PRIOR APPLICATION NUMBER: 60/084637
12 PRIOR FILING DATE: 1998-05-07
13 PRIOR APPLICATION NUMBER: 60/084639
14 PRIOR FILING DATE: 1998-05-07
15 PRIOR APPLICATION NUMBER: 60/084640
16 PRIOR FILING DATE: 1998-05-07
17 PRIOR APPLICATION NUMBER: 60/084598
18 PRIOR FILING DATE: 1998-05-07
19 PRIOR APPLICATION NUMBER: 60/084600
20 PRIOR FILING DATE: 1998-05-07
21 PRIOR APPLICATION NUMBER: 60/084627
22 PRIOR FILING DATE: 1998-05-07
23 PRIOR APPLICATION NUMBER: 60/084643
24 PRIOR FILING DATE: 1998-05-07
25 PRIOR APPLICATION NUMBER: 60/085339
26 PRIOR FILING DATE: 1998-05-13
27 PRIOR APPLICATION NUMBER: 60/085338
28 PRIOR FILING DATE: 1998-05-13
29 PRIOR APPLICATION NUMBER: 60/085323
30 PRIOR FILING DATE: 1998-05-13
31 PRIOR APPLICATION NUMBER: 60/085582
32 PRIOR FILING DATE: 1998-05-15
33 PRIOR APPLICATION NUMBER: 60/085700
34 PRIOR FILING DATE: 1998-05-15
35 PRIOR APPLICATION NUMBER: 60/085689
36 PRIOR FILING DATE: 1998-05-15
37 PRIOR APPLICATION NUMBER: 60/085579
38 PRIOR FILING DATE: 1998-05-15
39 PRIOR APPLICATION NUMBER: 60/085580
40 PRIOR FILING DATE: 1998-05-15
41 PRIOR APPLICATION NUMBER: 60/085573
42 PRIOR FILING DATE: 1998-05-15
43 PRIOR APPLICATION NUMBER: 60/085704
44 PRIOR FILING DATE: 1998-05-15
45 PRIOR APPLICATION NUMBER: 60/085697

```

| | | | | |
|-------------|-------|------------|-------|-------------|
| Query Match | 3.1%; | Score 128; | DB 9; | Length 432; |
|-------------|-------|------------|-------|-------------|

Matches 77; Conservative 52; Mismatches 140; Indels 116; Gaps 14;

| | | | |
|----|-----|---|------|
| QY | 379 | TEJENFNATVMASSDSIVLEFYGMQAVNAFICQSYIDAVVAKLKG--TSMILTRINCAWMSD | 436 |
| Dd | 67 | TADMTFHGISOAAHFVHVFEPFAWCGHCRLOPRTWMDLDDKINSMEDAKVUYAKDCATHSD | 1266 |
| QY | 437 | VCTKONATPEPIIIMYKKGPNPVSAYAGMLGTGDKLKFITQINRISYPNITISIOAEBYLS | 4966 |
| Dd | 127 | VCSAQGRGYTTLLFKRPGOEAVYQSGPROFQTLNNMLOTLNEEPP--TPREEVEPPSA | 1644 |
| QY | 497 | GELKPKDILLYSSVS-----VLGL-FSPPTMTAK | 5233 |
| Dd | 185 | PELMOGLIELTSASFELHVAHQGDHFIFKFAWCGHCKALAPTWEOQLAGLEHSEPTVATGK | 2444 |
| QY | 524 | EDFSE-----AGNYLKGVYITGIISEEDVLLSTKYAASLPALLARHTEGKTESIPLAS | 578 |
| Dd | 245 | VDCIQHRELCSGNQVRGY-----PTLLMFR--DGKVKDYQKKG | 280 |
| QY | 579 | THAODIVQIITDALLEFPEITVENLPSTYRLQPLILFSDGTVNPQYK-AILPLVKO | 637 |
| Dd | 281 | RDLESLREVEESQLORETGATETVPS--BAPVL-----AAPEADAKGTVALTEN | 330 |
| QY | 638 | KYLD-----SFPCCW--LNLKNTPVGRGLIRAYFD----- | 655 |
| Dd | 331 | NFDQITIAEGITFIKFAVPMWGHCKTALAPTWEEILSKKFFPGLAVCIYKLAEVQCTAERNICKS | 390 |

Oy 666 -PLPPLPLVL-----VNLSGGQ 683
Db 391 YSVRGYPTLLFRGGKKVSHSGGR 415

RESULT 10
US-09-978-697-90
: Sequence 90. Application US/09978697
: Patent No. US20020169284A1
: GENERAL INFORMATION:
: APPLICANT: Ashkenazi, Avi
: APPLICANT: Baker Kevin P.
: APPLICANT: Botstein, David
: APPLICANT: Desnovers, Luc
: APPLICANT: Eaton, Dan
: APPLICANT: Ferrara, Napoleon
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Fong, Sherman
: APPLICANT: Gao, Wei-Qiang
: APPLICANT: Gerber, Hanspeter
: APPLICANT: Gottard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Grimaldi, J. Christopher
: APPLICANT: Gurney, Austin L.
: APPLICANT: Hillan, Kenneth J.
: APPLICANT: Kijavijn, Ivar J.
: APPLICANT: Kuo, Sophia S.
: APPLICANT: Napier, Mary A.
: APPLICANT: Pan, James
: APPLICANT: Paonli, Nicholas F.
: APPLICANT: Roy, Margaret Ann
: APPLICANT: Shelton, David L.
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tunas, Daniel
: APPLICANT: Williams, P. Mickey
: APPLICANT: Wood, William I.
: TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
: TITLE OF INVENTION: Acids Encoding the Same
: FILE REFERENCE: P2630FIC27
: CURRENT APPLICATION NUMBER: US/09/978, 697
: CURRENT FILING DATE: 2001-10-16
: PRIOR APPLICATION NUMBER: 09/918585
: PRIOR FILING DATE: 2001-07-30
: PRIOR APPLICATION NUMBER: 60/062250
: PRIOR FILING DATE: 1997-10-17
: PRIOR APPLICATION NUMBER: 60/064249
: PRIOR FILING DATE: 1997-11-03
: PRIOR APPLICATION NUMBER: 60/065311
: PRIOR FILING DATE: 1997-11-13
: PRIOR APPLICATION NUMBER: 60/066364
: PRIOR FILING DATE: 1997-11-21
: PRIOR APPLICATION NUMBER: 60/077450
: PRIOR FILING DATE: 1998-03-10
: PRIOR APPLICATION NUMBER: 60/077632
: PRIOR FILING DATE: 1998-03-11
: PRIOR APPLICATION NUMBER: 60/077641
: PRIOR FILING DATE: 1998-03-11
: PRIOR APPLICATION NUMBER: 60/077649
: PRIOR FILING DATE: 1998-03-11
: PRIOR APPLICATION NUMBER: 60/077791
: PRIOR FILING DATE: 1998-03-12
: PRIOR APPLICATION NUMBER: 60/078004
: PRIOR FILING DATE: 1998-03-13
: PRIOR APPLICATION NUMBER: 60/078886
: PRIOR FILING DATE: 1998-03-20
: PRIOR APPLICATION NUMBER: 60/078936
: PRIOR FILING DATE: 1998-03-20
: PRIOR APPLICATION NUMBER: 60/078910
: PRIOR FILING DATE: 1998-03-20
: PRIOR APPLICATION NUMBER: 60/078939
: PRIOR FILING DATE: 1998-03-20

: PRIOR APPLICATION NUMBER: 60/079294
: PRIOR FILING DATE: 1998-03-25
: PRIOR APPLICATION NUMBER: 60/079656
: PRIOR FILING DATE: 1998-03-26
: PRIOR APPLICATION NUMBER: 60/079664
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/079689
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/079663
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/079728
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/079786
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/079920
: PRIOR FILING DATE: 1998-03-30
: PRIOR APPLICATION NUMBER: 60/079923
: PRIOR FILING DATE: 1998-03-30
: PRIOR APPLICATION NUMBER: 60/080105
: PRIOR FILING DATE: 1998-03-31
: PRIOR APPLICATION NUMBER: 60/080107
: PRIOR FILING DATE: 1998-03-31
: PRIOR APPLICATION NUMBER: 60/080165
: PRIOR FILING DATE: 1998-03-31
: PRIOR APPLICATION NUMBER: 60/080194
: PRIOR FILING DATE: 1998-03-31
: PRIOR APPLICATION NUMBER: 60/080327
: PRIOR FILING DATE: 1998-04-01
: PRIOR APPLICATION NUMBER: 60/080328
: PRIOR FILING DATE: 1998-04-01
: PRIOR APPLICATION NUMBER: 60/080333
: PRIOR FILING DATE: 1998-04-01
: PRIOR APPLICATION NUMBER: 60/080334
: PRIOR FILING DATE: 1998-04-01
: PRIOR APPLICATION NUMBER: 60/081070
: PRIOR FILING DATE: 1998-04-08
: PRIOR APPLICATION NUMBER: 60/081049
: PRIOR FILING DATE: 1998-04-08
: PRIOR APPLICATION NUMBER: 60/081071
: PRIOR FILING DATE: 1998-04-08
: PRIOR APPLICATION NUMBER: 60/081195
: PRIOR FILING DATE: 1998-04-08
: PRIOR APPLICATION NUMBER: 60/081203
: PRIOR FILING DATE: 1998-04-09
: PRIOR APPLICATION NUMBER: 60/081229
: PRIOR FILING DATE: 1998-04-09
: PRIOR APPLICATION NUMBER: 60/081955
: PRIOR FILING DATE: 1998-04-15
: PRIOR APPLICATION NUMBER: 60/081817
: PRIOR FILING DATE: 1998-04-15
: PRIOR APPLICATION NUMBER: 60/081819
: PRIOR FILING DATE: 1998-04-15
: PRIOR APPLICATION NUMBER: 60/081952
: PRIOR FILING DATE: 1998-04-15
: PRIOR APPLICATION NUMBER: 60/081838
: PRIOR FILING DATE: 1998-04-15
: PRIOR APPLICATION NUMBER: 60/082568
: PRIOR FILING DATE: 1998-04-21
: PRIOR APPLICATION NUMBER: 60/082569
: PRIOR FILING DATE: 1998-04-21
: PRIOR APPLICATION NUMBER: 60/082704
: PRIOR FILING DATE: 1998-04-22
: PRIOR APPLICATION NUMBER: 60/082804
: PRIOR FILING DATE: 1998-04-22
: PRIOR APPLICATION NUMBER: 60/082700
: PRIOR FILING DATE: 1998-04-22
: PRIOR APPLICATION NUMBER: 60/082797
: PRIOR FILING DATE: 1998-04-22
: PRIOR APPLICATION NUMBER: 60/082796
: PRIOR FILING DATE: 1998-04-23
: PRIOR APPLICATION NUMBER: 60/083336
: PRIOR FILING DATE: 1998-04-27
: PRIOR APPLICATION NUMBER: 60/083322

```

1 PRIOR FILING DATE: 1998-04-28
2 PRIOR APPLICATION NUMBER: 60/083392
3 PRIOR FILING DATE: 1998-04-29
4 PRIOR APPLICATION NUMBER: 60/083495
5 PRIOR FILING DATE: 1998-04-29
6 PRIOR APPLICATION NUMBER: 60/083496
7 PRIOR FILING DATE: 1998-04-29
8 PRIOR APPLICATION NUMBER: 60/083499
9 PRIOR FILING DATE: 1998-04-29
10 PRIOR APPLICATION NUMBER: 60/083445
11 PRIOR FILING DATE: 1998-04-29
12 PRIOR APPLICATION NUMBER: 60/083554
13 PRIOR FILING DATE: 1998-04-29
14 PRIOR APPLICATION NUMBER: 60/083558
15 PRIOR FILING DATE: 1998-04-29
16 PRIOR APPLICATION NUMBER: 60/083559
17 PRIOR FILING DATE: 1998-04-29
18 PRIOR APPLICATION NUMBER: 60/083500
19 PRIOR FILING DATE: 1998-04-29
20 PRIOR APPLICATION NUMBER: 60/083742
21 PRIOR FILING DATE: 1998-04-30
22 PRIOR APPLICATION NUMBER: 60/083366
23 PRIOR FILING DATE: 1998-05-05
24 PRIOR APPLICATION NUMBER: 60/084414
25 PRIOR FILING DATE: 1998-05-06
26 PRIOR APPLICATION NUMBER: 60/084441
27 PRIOR FILING DATE: 1998-05-06
28 PRIOR APPLICATION NUMBER: 60/084637
29 PRIOR FILING DATE: 1998-05-07
30 PRIOR APPLICATION NUMBER: 60/084639
31 PRIOR FILING DATE: 1998-05-07
32 PRIOR APPLICATION NUMBER: 60/084640
33 PRIOR FILING DATE: 1998-05-07
34 PRIOR APPLICATION NUMBER: 60/084598
35 PRIOR FILING DATE: 1998-05-07
36 PRIOR APPLICATION NUMBER: 60/084600
37 PRIOR FILING DATE: 1998-5-07
38 PRIOR APPLICATION NUMBER: 60/084627
39 PRIOR FILING DATE: 1998-05-07
40 PRIOR APPLICATION NUMBER: 60/084643
41 PRIOR FILING DATE: 1998-05-07
42 PRIOR APPLICATION NUMBER: 60/085339
43 PRIOR FILING DATE: 1998-05-13
44 PRIOR APPLICATION NUMBER: 60/085338
45 PRIOR FILING DATE: 1998-05-13
46 PRIOR APPLICATION NUMBER: 60/085323
47 PRIOR FILING DATE: 1998-05-13
48 PRIOR APPLICATION NUMBER: 60/085582
49 PRIOR FILING DATE: 1998-05-15
50 PRIOR APPLICATION NUMBER: 60/085700
51 PRIOR FILING DATE: 1998-05-15
52 PRIOR APPLICATION NUMBER: 60/085689
53 PRIOR FILING DATE: 1998-05-15
54 PRIOR APPLICATION NUMBER: 60/085579
55 PRIOR FILING DATE: 1998-05-15
56 PRIOR APPLICATION NUMBER: 60/085580
57 PRIOR FILING DATE: 1998-05-15
58 PRIOR APPLICATION NUMBER: 60/085573
59 PRIOR FILING DATE: 1998-05-15
60 PRIOR APPLICATION NUMBER: 60/085704
61 PRIOR FILING DATE: 1998-05-15
62 PRIOR APPLICATION NUMBER: 60/085697

```

| | | | | |
|-----------------------|------------------|------------------|-------------|-------------|
| Query Match | 3.1%; | Score 128; | DB 9; | Length 432; |
| Best Local Similarity | 20.0%; | Pred. No. 0.015; | | |
| Matches 77; | Conservative 52; | Mismatches 140; | Indels 116; | Gaps 14 |

[illegible]

| | | | | |
|----|-----|--|-------------------------|------|
| Db | 127 | VCSAQGVGRVPTLKLFRKGQDAVYKQGRDQGLTEMMILQTLNBEV- | -TPREVEBP | 1844 |
| Qy | 497 | GELIKDILLYSVS- | -VLGL-FSPYMKTA | 523 |
| Db | 185 | PELKGGILETASNEELHVAOGDHIFKFFAWGCHKALAPTWEOQLA | LGHESTVYIGK | 2444 |
| Qy | 524 | EDFSE-----AGNYLKGVIITGISDEDVLLSTKYAASLPALLAR | HTEGKTESIPLAS | 578 |
| Db | 245 | VDCTGHVELGSGNOVGRV- | PTLLMFR-DGKRVQYK | 280 |
| Qy | 579 | THAADIIVITDALLEMPETIVENLSYFRLOKPLLIESDGINVQYKK | -ALITLVQ | 637 |
| Db | 281 | RDLSELRVEVSQLORFETGATETVTS--- | EAPVL---AAPEADKGTVALTEN | 330 |
| Qy | 638 | KYLD-----SFTPCW--LNLKMT | PVGRGLIRAYFD----- | 665 |
| Db | 331 | NFDDIAGETIFIKFYAPWGCCKTAPLTMWELSKKEPGLAGVAKIA | LEVDTAERNICSK | 350 |
| Qy | 666 | -PLPPLVLVL-----VNLHSGQ | 683 | |
| Db | 391 | YSVRGYPITLLFRGKKVSEHSGR | 415 | |

RESULT 11
 US-09-978-192A-90
 Sequence 90: Application US/09978192A
 Patent No. US2002017753A1
 GENERAL INFORMATION:
 APPLICANT: Ashkenazi, Avi
 APPLICANT: Baker Kevin P.
 APPLICANT: Botstein, David
 APPLICANT: Desnovers, Luc
 APPLICANT: Eaton, Dan
 APPLICANT: Ferrara, Napoleon
 APPLICANT: Filvaroff, Ellen
 APPLICANT: Fong, Sherman
 APPLICANT: Gao, Wei-Qiang
 APPLICANT: Gerber, Hanspeter
 APPLICANT: Gerritsen, Mary E.
 APPLICANT: Goddard, Audrey
 APPLICANT: Godowski, Paul J.
 APPLICANT: Grimaldi, J. Christopher
 APPLICANT: Gurney, Austin L.
 APPLICANT: Hillan, Kenneth J.
 APPLICANT: Kijavrin, Ivar J.
 APPLICANT: Kuo, Sophia S.
 APPLICANT: Napier, Mary A.
 APPLICANT: Paoni, Nicholas F.
 APPLICANT: Roy, Margaret Ann
 APPLICANT: Shelton, David L.
 APPLICANT: Stewart, Timothy A.
 APPLICANT: Tumas, Daniel
 APPLICANT: Williams, P. Mickey
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
 TITLE OR INVENTION: Acids Encoding the Same
 FILE REFERENCE: P2630PIC9
 CURRENT APPLICATION NUMBER: US/09/978,192A
 CURRENT FILING DATE: 2001-10-15
 PRIOR APPLICATION NUMBER: 09/918585
 PRIOR FILING DATE: 2001-07-30
 PRIOR APPLICATION NUMBER: 60/062250
 PRIOR FILING DATE: 1997-10-17
 PRIOR APPLICATION NUMBER: 60/064249
 PRIOR FILING DATE: 1997-11-03
 PRIOR APPLICATION NUMBER: 60/065311
 PRIOR FILING DATE: 1997-11-13
 PRIOR APPLICATION NUMBER: 60/066364
 PRIOR FILING DATE: 1997-11-21
 PRIOR APPLICATION NUMBER: 60/077450
 PRIOR FILING DATE: 1998-03-10
 PRIOR APPLICATION NUMBER: 60/077632


```
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085573
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697

Query Match      3.1%; Score 128; DB 9; Length 432;
Best Local Similarity 20.0%; Pred. No. 0.015;
Matches 77; Conservative 52; Mismatches 140; Indels 116; Gaps 14;

QY 379 TEETFNATVVASDSIVFYFGMAQVSNAPFIQSYIDVAVKLG--TSMMLTRINCADMSD 436
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 67 TADMFTHGIGSAHFWEFFAFWCGHCGRLOPTWMDLDGKNSMEDAKVYAKVDCSTHSD 126
QY 437 VCTKQNTPEPTIKMYKKGPNVSYAGMLGSKDLEKFIQINRISYPNITISIOAEYLS 496
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 127 VCSAQGVKRGPTLKLFRPGDEAKYQGRPDFOILENNMLOTLNEEPV--TPPEVEPPSA 184
QY 497 GELYKDLILYSVS-----VLGL-FSPPTMTAK 523
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 185 PELKQGLYELASNFELHVAQGDHFIFAFWCGHCALAPTWEOQLALGLEHSEYVIGK 244
QY 524 EDSE-----AGNYLKGYVTGTIYSEEDVLLSTKYASLPALLNHTGKTESIPLAS 578
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 245 VDCIOHYELCSGNQVRGY-----PTLLWFR--DGKKVDQYK 280
QY 579 THADQIVQITDALLEFPEITVENLPSYFLQKPLILFESDGTVPQYK-AIITLVKQ 637
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 281 RQLESLEREYVESQLORETCATETVPS---EAPVL-----AAPEAKGVIALTEN 330
QY 638 KYLD-----SFTPCN--LNKNTPVYRGILRAYFD----- 665
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 331 NFDDTIAEGITFTKFPYAPWCGHCKTLAPTWELSKKEFPGLAGVIAEVCTAERNICSK 390
QY 666 -PLPPLPLVL-----VNLHSGQ 683
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 391 YSVRGYPTLLFRGKKVSEHSGR 415

RESULT 12
US-09-999-832A-90
; Sequence 90, Application US/09999832A
; Publication No. US20020192706A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Baker Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gertlisen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Kuo, Sophia S.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James;
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Shelton, David L.
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
```

```
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2630PIC3
; CURRENT APPLICATION NUMBER: US/09/999,832A
; CURRENT FILING DATE: 2001-10-24
; PRIOR APPLICATION NUMBER: 09/918585
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064249
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077641
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078004
; PRIOR FILING DATE: 1998-03-13
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078936
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079920
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/079923
; PRIOR FILING DATE: 1998-03-30
; PRIOR APPLICATION NUMBER: 60/080105
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080328
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080334
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081071
```

| Query Match | 3.1% | Score 128: | DB 9: | Length 432: |
|---|---|------------------|-------------|-------------|
| Best Local Similarity | 20.08: | Pred. No. 0.015: | | |
| Matches 77: | Conservative 52: | Mismatches 140: | Indels 116: | Gaps 14: |
| YQ 379 TEEFNATVMA SDSIVLEFAGQAVSMAFLOSVIDAVAKLKG--NSTMLTRINCADMSD 436 | : : : : : : : : : : : : : : : : : : : : : : : : : | | | |
| Db 67 TADFTHTGIGSAHVFWMFFAPWCGHCORLQPTWNDLGGKYSMEDEAKYVAVKDVCTAHS 126 | : : : : : : : : : : : : : : : : : : : : : : : : : | | | |
| YQ 437 VCTQNVTEPPTITMYKKGKGVNPSYAGMLGTKDLKFTQLNRSYVNVITSIQEAEYLS 496 | : : : : : : : : : : : : : : : : : : : : : : : : : | | | |
| Db 127 VCSAQGVARGVPTTLKFKRGQEA VKKQGRDQTLNMMQLQTLNEBPV--TPEPEVEPPSA 184 | : : : : : : : : : : : : : : : : : : : : : : : : : | | | |
| YQ 497 GELYKDLILYSVS-----VGLF-FSPIMKTAK 523 | : : : : : : : : : : : : : : : : : : : : : : : : : | | | |
| Db 185 PELQGLYEELSSASFELHVAQGDHIFKEFAPWCGHCKALAPTEOALGLHESETVKG 244 | : : : : : : : : : : : : : : : : : : : : : : : : : | | | |
| YQ 524 EDGSE-----AGNVLKGVIITIGYSEEDVLLSTFYASLPAALLARHTEGIESIPLAS 578 | : : : : : : : : : : : : : : : : : : : : : : : : : | | | |
| Db 245 VDCQHYELCGSGNOVRG-----PTLLMFR--DGKKVDQYK 280 | : : : : : : : : : : : : : : : : : : : : : : : : : | | | |
| YQ 579 THAADIQIITDALLEMPETIVENTLPSYFRLOKPLILFSDGVYVPOYKKAITLIVKO 637 | : : : : : : : : : : : : : : : : : : : : : : : : : | | | |
| Db 281 RDLESRLREYVESQLORTGTGATEVTPS-----EAYL-----AAEPADKQVIALTEN 330 | : : : : : : : : : : : : : : : : : : : : : : : : : | | | |
| YQ 638 KYLD-----SFTPCW-LNEKNTPVGRIIRAYED----- 665 | : : : : : : : : : : : : : : : : : : : : : : : : : | | | |
| Db 331 NFDFTIAEGITFIKFYAPWCGHCKTLATPTWBELSKKEPGLAGVKAIAVDCTAERNICSK 390 | : : : : : : : : : : : : : : : : : : : : : : : : : | | | |
| YQ 666 -PLRPLLVL-----VNLSCGQ 683 | : : : : : : : : : : : : : : : : : : : : : : : : : | | | |
| Db 391 YSVRGYPTLLLFRRGKKVSEHSGR 415 | : : : : : : : : : : : : : : : : : : : : : : : : : | | | |

APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Kuo, Sophia S.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James;
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Shelton, David L.
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2630PIC7
CURRENT APPLICATION NUMBER: US/09/978,109
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/064249
PRIOR FILING DATE: 1997-11-03
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066364
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077641
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: 60/078004
PRIOR FILING DATE: 1998-03-13
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078936
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079663
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079920
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/079923
PRIOR FILING DATE: 1998-03-30
PRIOR APPLICATION NUMBER: 60/080105
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31

PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080328
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080334
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081071
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081203
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081229
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081952
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082700
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082796
PRIOR FILING DATE: 1998-04-23
PRIOR APPLICATION NUMBER: 60/083336
PRIOR FILING DATE: 1998-04-27
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083392
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083554
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083558
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083500
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083742
PRIOR FILING DATE: 1998-04-30
PRIOR APPLICATION NUMBER: 60/084366

```

1 PRIOR FILING DATE: 1998-05-05
2 PRIOR APPLICATION NUMBER: 60/084414
3 PRIOR FILING DATE: 1998-05-06
4 PRIOR APPLICATION NUMBER: 60/084441
5 PRIOR FILING DATE: 1998-05-06
6 PRIOR APPLICATION NUMBER: 60/084637
7 PRIOR FILING DATE: 1998-05-07
8 PRIOR APPLICATION NUMBER: 60/084639
9 PRIOR FILING DATE: 1998-05-07
10 PRIOR APPLICATION NUMBER: 60/084640
11 PRIOR FILING DATE: 1998-05-07
12 PRIOR APPLICATION NUMBER: 60/084598
13 PRIOR FILING DATE: 1998-05-07
14 PRIOR APPLICATION NUMBER: 60/084600
15 PRIOR FILING DATE: 1998-5-07
16 PRIOR APPLICATION NUMBER: 60/084627
17 PRIOR FILING DATE: 1998-05-07
18 PRIOR APPLICATION NUMBER: 60/084643
19 PRIOR FILING DATE: 1998-05-07
20 PRIOR APPLICATION NUMBER: 60/085339
21 PRIOR FILING DATE: 1998-05-13
22 PRIOR APPLICATION NUMBER: 60/085338
23 PRIOR FILING DATE: 1998-05-13
24 PRIOR APPLICATION NUMBER: 60/085323
25 PRIOR FILING DATE: 1998-05-13
26 PRIOR APPLICATION NUMBER: 60/085582
27 PRIOR FILING DATE: 1998-05-15
28 PRIOR APPLICATION NUMBER: 60/085700
29 PRIOR FILING DATE: 1998-05-15
30 PRIOR APPLICATION NUMBER: 60/085689
31 PRIOR FILING DATE: 1998-05-15
32 PRIOR APPLICATION NUMBER: 60/085579
33 PRIOR FILING DATE: 1998-05-15
34 PRIOR APPLICATION NUMBER: 60/085580
35 PRIOR FILING DATE: 1998-05-15
36 PRIOR APPLICATION NUMBER: 60/085573
37 PRIOR FILING DATE: 1998-05-15
38 PRIOR APPLICATION NUMBER: 60/085704
39 PRIOR FILING DATE: 1998-05-15
40 PRIOR APPLICATION NUMBER: 60/085697

```

```

US-10-174-590-74
RESULT 14
; Sequence 74, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C42
; CURRENT APPLICATION NUMBER: US/10/174,590
; CURRENT FILING DATE: 2002-06-18
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 74
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Homo Saplen
; US-10-174-590-74

```

| | | | | |
|-----------------------|------------------|------------------|-------------|-------------|
| Query Match | 3.1%; | Score 128; | DB 9; | Length 432; |
| Best Local Similarity | 20.0%; | Pred. No. 0.015; | | |
| Matches 77; | Conservative 52; | Mismatches 140; | Indels 116; | Gaps 14 |

```

Db      281 RDESLREEVESQLOFETGATETVPS-----EAPVL-----AAPEAARGFVALTEN 330
QY      638 KYLD-----SFFPCW--LNLKNTPGYGRGLRAYD----- 665
Db      331 NEDFTLIEGTIFKFAVPMCGHCCKTAPLPTWELSKKEFPGLAGVIAEVDCTAERNICSK 390
QY      666 -PLPPLPLVL-----VNLHSGGQ 683
           :   |   |   :   |   |
Db      391 YSVRGYPTLLLFROGKRAVSEHSGGR 415

```

```

RESULT 15
US-10-176-758-74
; Sequence 74, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.

```

```

: APPLICANT: Gurney,Austin L.
: APPLICANT: Pan,James
: APPLICANT: Smith,Victoria
: APPLICANT: Watanabe,Colin K.
: APPLICANT: Wood,William I.
: APPLICANT: Zhang,Zemin
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEOTIC
: TITLE OF INVENTION: ACIDS ENCODING THE SAME
: FILE REFERENCE: P3430R1C104
: CURRENT APPLICATION NUMBER: US/10/176,758
: PRIORITY FILING DATE: 2002-06-21
: Prior Application removed - See File Wrapper or Palm
: NUMBER OF SEQ ID NOS: 612
: SEQ ID NO 74
: LENGTH: 432
: TYPE: PRP
: ORGANISM: Homo Sapien
: US-10-176-758-74

```

| | | | | |
|-----------------------|------------------|------------------|-------------|-------------|
| Query Match | 3.18; | Score 128; | DB 9; | Length 432; |
| Best Local Similarity | 20.08; | Pred. No. 0.015; | | |
| Matches 77; | Conservative 52; | Mismatches 140; | Indels 116; | Gaps 14 |

[illegible]

Search completed: March 13, 2003, 17:18:06
Job time : 21 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:11:47 ; Search time 21 Seconds
(without alignments)
3694.309 Million cell updates/sec

Title: US-09-847-046-2

Perfect score: 4142
Sequence: 1 MFGGFNVFRVIGSFVIMCIF.....FKEAEKSFRRDKELGCSKVN 807

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

- 1: PIR_73:*
- 2: PIR1:*
- 3: PIR2:*
- 4: PIR3:*
- 5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 185 | 4.5 | 505 | 2 JC2291 | protein disulfide- |
| 2 | 182 | 4.4 | 546 | 2 P96552 | hypothetical prote |
| 3 | 172.5 | 4.2 | 664 | 2 S44756 | probable protein d |
| 4 | 169.5 | 4.1 | 482 | 2 S34275 | protein disulfide- |
| 5 | 169.5 | 4.1 | 505 | 2 JC5704 | protein disulfide- |
| 6 | 169.5 | 4.1 | 505 | 2 S68363 | protein disulfide- |
| 7 | 169.5 | 4.1 | 515 | 2 S57942 | protein disulfide- |
| 8 | 167 | 4.0 | 488 | 1 JC2385 | protein disulfide- |
| 9 | 166.5 | 4.0 | 505 | 1 S55507 | protein disulfide- |
| 10 | 166.5 | 4.0 | 505 | 2 S63994 | protein disulfide- |
| 11 | 154.5 | 3.7 | 522 | 1 ISBY6S | protein disulfide- |
| 12 | 153 | 3.7 | 504 | 1 A28807 | protein disulfide- |
| 13 | 152 | 3.7 | 504 | 2 S41661 | protein disulfide- |
| 14 | 150.5 | 3.6 | 517 | 2 JC7623 | protein disulfide- |
| 15 | 148 | 3.6 | 566 | 2 T06724 | protein disulfide- |
| 16 | 147 | 3.5 | 497 | 1 A32820 | protein disulfide- |
| 17 | 146.5 | 3.5 | 492 | 2 T38093 | probable protein d |
| 18 | 144.5 | 3.5 | 359 | 2 T03644 | probable protein d |
| 19 | 144 | 3.5 | 643 | 1 S32476 | protein disulfide- |
| 20 | 143.5 | 3.5 | 2471 | 2 T42977 | large tegument pro |
| 21 | 138.5 | 3.3 | 645 | 1 A23723 | protein disulfide- |
| 22 | 137 | 3.3 | 769 | 2 S54525 | mismatch repair pr |
| 23 | 137 | 3.3 | 2493 | 2 S45734 | probable membrane |
| 24 | 136 | 3.3 | 508 | 1 ISRTSS | protein disulfide- |
| 25 | 136 | 3.3 | 509 | 1 A38362 | protein disulfide- |
| 26 | 135 | 3.3 | 508 | 1 ISHUS5 | protein disulfide- |
| 27 | 135 | 3.3 | 510 | 1 ISBOSS | protein disulfide- |
| 28 | 134.5 | 3.2 | 440 | 2 T01115 | probable protein d |
| 29 | 134.5 | 3.2 | 517 | 2 A44483 | protein disulfide |

| | | | | | |
|----|-------|-----|------|----------|--------------------|
| 30 | 133.5 | 3.2 | 584 | 2 S06318 | endoplasmic reticu |
| 31 | 133.5 | 3.2 | 638 | 1 ISMSER | protein disulfide- |
| 32 | 133 | 3.2 | 361 | 2 T00437 | probable protein d |
| 33 | 133 | 3.2 | 572 | 2 T32636 | hypothetical prote |
| 34 | 132 | 3.2 | 2829 | 2 A42771 | reticulocyte-bindi |
| 35 | 131.5 | 3.2 | 364 | 2 T09614 | probable protein d |
| 36 | 131.5 | 3.2 | 574 | 2 T25887 | hypothetical prote |
| 37 | 131 | 3.2 | 1088 | 1 P1XRSR | inner layer protei |
| 38 | 130.5 | 3.2 | 363 | 2 T37630 | protein disulfide- |
| 39 | 130.5 | 3.2 | 485 | 2 S71863 | protein disulfide- |
| 40 | 128.5 | 3.1 | 2149 | 2 T47655 | hypothetical prote |
| 41 | 128 | 3.1 | 977 | 2 E86349 | hypothetical prote |
| 42 | 127.5 | 3.1 | 488 | 2 T23055 | hypothetical prote |
| 43 | 127.5 | 3.1 | 489 | 2 S68280 | protein disulfide- |
| 44 | 127.5 | 3.1 | 509 | 1 ISMSSS | protein disulfide- |
| 45 | 127 | 3.1 | 443 | 2 D86183 | hypothetical prote |

ALIGNMENTS

RESULT 1

JC2291
protein disulfide-isomerase (EC 5.3.4.1) precursor - Humicola insolens

C:Species: Humicola insolens
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 28-May-1999

C:Accession: JC2291; PC2200

R:Kajino, T.; Sarai, K.; Imada, T.; Idekoba, C.; Asami, O.; Yamada, Y.; Hirai, M.; U

Biosci. Biotechnol. Biochem. 58, 1424-1429, 1994

A:Title: Molecular cloning of a fungal cDNA encoding protein disulfide isomerase.

A:Reference number: JC2291; MUID:94369094; PMID:7765273

A:Accession: JC2291

A:Molecule type: mRNA

A:Residues: 1-505 <KAD>

A:Cross-references: GB:S74296; NID:9712822; PIDN:AC60578.1; PID:9712823

A:Accession: PC2200

A:Molecule type: protein

A:Residues: 21-46;55-66;77-110;115-124;171-187;207-212;330-339;353-374;395-404;416-43

C:Comment: This enzyme catalyzes thiol-protein-disulfide interchange in vitro, with v

C:Superfamily: protein disulfide-isomerase; thioredoxin homology

C:Keywords: glycoprotein; intramolecular oxidoreductase; isomerase

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-505/Product: protein disulfide-isomerase #status predicted <MAT>

F:29-112/Domain: thioredoxin homology <TX1>

F:363-450/Domain: thioredoxin homology <TX2>

F:105/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 4.5%; Score 185; DB 2; Length 505;
Best Local Similarity 21.9%; Pred. No. 0.00023;
Matches 112; Conservative 79; Mismatches 213; Indels 108; Gaps 22;

| | | | |
|----|-----|---|-----|
| QY | 15 | VIMCFYMYPTVNSLELSPOKIFSTLQPGLELNANVRPLQDYGISVAKVNCVKEISRY | 74 |
| DB | 40 | LVLAEPFAPWCCHCKALAP-----EYEAATTLKKNIKLAKVDT----- | 80 |
| QY | 75 | CGKEKDLKMAVLFKGNILRER-----PTGTLFDVNAVIVAHVFALL--FSEVAYITN | 125 |
| DB | 81 | --EEDFLCQGHVEGYPPLTKVNRGLDWSFYGGQKAAITSTYMKQSLPANSVETK-DN | 137 |
| QY | 126 | LEDLQNIENALKKANIIFSYRAIGIPRHRAVMAGFYVGTGYFVLTETALLSISGS | 185 |
| DB | 138 | LEEFKADKA-----VVAAYVDASDKASSEVFQVAEKLRDNYDPFGSSSDAALAEAGV | 191 |
| QY | 166 | EEVEFAHLFFHCRLVLDLTOGCRRTLMQPLTTINIHIFITMKAPLITFEVAEDPOVS | 245 |
| DB | 192 | KAP-----ATVLLKDEDEGKAVSEKEVEAIEFAATGATPLIGETG--PEYTS | 239 |
| QY | 246 | TYHLQGLPLVEIVSQQA-----TTEADRTAEW--VAMRLIG--KAGVILLLR | 290 |
| DB | 240 | D-YMAGIPLAVITFAETBEERKELSDKLPIAEARGVINFTIDAKAFGAIAGNLNKT | 298 |
| QY | 291 | DSLEV-----NIPQDANVVFKAEEGVVPEFLVLDVLLIISVENVNMHIEIQEEDNDM | 346 |

Db 299 DKFPAI0EVAKNQKPFPEQKE-ITFEAIKAFVDDFVAGKIEPSIKSPIPEKO---- 333
Qy 347 EGGPDIVODEVAETVFRDRKRRLPLELVELTEETFNATVMASSIVLEFYAGQVMSMA 406
Db 354 EGGVTVVAAKNNEIYLDTRK-----DVLEETFAWCGCHKA 390
Qy 407 FLOSYIDVA---VKLGSTMLLRINCADMSVCTKQNTPEPIIKMYKGE--NPVSY 461
Db 391 LAFKYEELGALYAKSEFKDPRVIKVD-ATANDV--PDEI0GFPTIKLYPAGAKGQPVTV 447
Qy 462 AGMLGTCKDLKFKTQLRNRSYPVNITSIOEAE 493
Db 448 SGSRVIEDLIKFTAEAG-KYKAIS--EDAE 476

RESULT 2

P96562
hypothetical protein F19K6.17 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: F96562
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Malt, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MIMD:21016719; PMID:11130712
A:Accession: F96562
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-546 <STO>
A:Cross-references: GB:AE005173; NID:g10645454; PIDN:AA21570.1; GSPDB:GN00141
A:Gene: F19K6.17
A:Map position: 1

Query Match 4.4%; Score 182; DB 2; Length 546;

Best Local Similarity 20.9%; Pred. No. 0.00041;
Matches 111; Conservative 93; Mismatches 180; Indels 148; Gaps 25;

Qy 317 LVLDHVDLIISHVENMHEIEQEDNDME--CPDIDVQDD-----EVAETVFRDRKR 368
Db 19 LILLSTIIIAVSSPDSNENEPGDSDDLQDLAYDEQLQDEPRPOQSAETV--SKAQ 76
Qy 369 KLPLELVELTEETFNATVMASSIVLEFYAGQVMSMAFLQSYIDVAVKLGFT-STMLLT 427
Db 77 RYVLELNGDYTKRVIDGNEFV--NVIGYAPMCARSALMPRAEATATLKEIGSVLMA 133
Qy 428 RINCADMSVCTKQNTPEPIIKMYKGENPVSYAGMLGKDLKFKTQLRNRSYPVNITS 487
Db 134 KIDGDRYSIASLELTKGPTLLFLVNGTS-LTYNGSSASDVIIVOKKTGAPITTLNT 192
Qy 488 IOEAEYELSGELYKDLILVSSVGLGF-----SPTMKTAKEP-----FSE----- 528
Db 193 VDEAPRFL--DKYHTF-----VLGLFEKFESEHNEFYKAAKSDEIOFETROSDVA 243
Qy 529 -----AGNYLKGY-----VITGIYSEEDVL--LSTKY----- 554
Db 244 KLLFPLKSNNVFGLVPEAEERYTVYDGSYKMEKILFELGSKNKPFLFKLTETNTVWVY 303
Qy 555 ---AASLPALLARHTGKIESIPLASTHAODIVQIITDALLEMPETIYENLPSTYRLOK 612
Db 304 SSPVKLQVWLFKADDFOKLAQPL-----EDIAKFKSKMFTYVITITEN-----LAM 352
Qy 613 PLLILFSDGVNPOYKALLTLVKOKYLDSPFCMILNKN--TPVGRGILRAYF--DPLP 668
Db 353 PLLILFGLTAGKNTVVAADNNLNLSKYLLSDPSPNSIEFCCGLAHGIVYSRTYRSEPV 412

Qy 669 PLPLVLVNLHSGGVFAFPSSQAIT--EENLVL-----WLKLEA-----G 708
Db 413 DNEIASIVTV--VGKTF---DGLVINSRENVLLHEHTPVCNCEALSKOIERLAKHFKG 466
Qy 709 LENHIITLPAQEWKPLPAYDFLSMIDAAATSOGTGRKVRPCMKETGVQENDK 760
Db 467 FENLV-----FARIDASANE-----HTKLQYDDK 490

RESULT 3

S44756
probable protein disulfide-isomerase (EC 5.3.4.1) - Caenorhabditis elegans
N:Alternate names: C14B9.2 protein
C:Species: Caenorhabditis elegans
C:Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 16-Jul-1999
C:Accession: S44756
R:Favell, A.D.
submitted to the EMBL Data Library, May 1993
A:Description: Sequence of the C. elegans cosmid C14B9.
A:Reference number: S44617
A:Accession: S44756
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-664 <FAV>
A:Cross-references: EMBL:L15188; NID:g289640; PIDN:AAA27952.1; PID:g289643
C:Gene: C14B9.2
A:Authors: 43/2; 110/2; 631/3
C:Superfamily: protein disulfide-isomerase; thioredoxin homology
C:Keywords: intramolecular oxidoreductase; isomerase
F:90-170/Domain: thioredoxin homology <TX1>
F:201-285/Domain: thioredoxin homology <TX2>
F:553-640/Domain: thioredoxin homology <TX3>

Query Match 4.2%; Score 172.5; DB 2; Length 664;

Best Local Similarity 19.4%; Pred. No. 0.0024;
Matches 87; Conservative 90; Mismatches 175; Indels 97; Gaps 17;

Qy 345 DMEGP-DIDVQDE--VAETVFR--DRKRKPLELVELTEETFNATVMASSIVL-FYA 398
Db 160 DKGPNVDYDGRDEAGIVEVESRVDPNKKPPEEVYTLTENFDIFSNLVLVEFYA 219
Qy 399 GMAVSMALQSYIDVAVKLGFT-STMLLRINCADMSVCTKQNTPEPIIKMYKGEN 457
Db 220 PWCGHCKKLAPERYEKAQKILKAGSKVKLGKVDATTEKDGTVYGVSPYTKIIRNGR 279
Qy 458 PVSYAGMLGKDLKFKTQLRNRSYPVNITSIOEAEYELSGELYKDLILVSSVGLGFSP 517
Db 280 -FPIYNGPREAAGIITMTQSKRAAKKLPKDKVERFMS---KD-----DVITIGFPA 329
Qy 518 TMTAKEDPSEAGNYLKGYITGIYSEE-----DVLLSTKYAASLPALLAR 565
Db 330 EDSTAEARSDSDEMRLREERKTGHTSDPAAFKMDAKPRDITIF-----YPSLFHSK 382
Qy 566 HTGKIESIPLASTHAODIVQIITDALLEMPETIYENLPSTYRLOKPLILFSDGVN 625
Db 383 F-EPKSRITYNKAATSEDLAEPFRESAPLVGCKTKKNAATR--TKRPLVVVYNAADFVS 440
Qy 626 QYKKA-----LTLVKOKYLDSPFCMILNKNTPVGRGILRAYDPLPPLPLVLVNL 678
Db 441 QYREGSEYRMSKYLNAQYIQRDKY----- 466
Qy 679 HSGGVFAFPSSQAIT--EENLVL-----WLKLEA-----G 708
Db 467 -----FAVADEEFAKE-----LEELGLGDSGLEHNVVYFGDGKKYPMNDEFGELD 515
Qy 736 ---AATSOGRTRKVRPCMKETGVQENDK 760
Db 516 ENLEAFMKQISSGKAKAHVKSAPAPKDK 544

RESULT 4

S34275
protein disulfide-isomerase homolog precursor - fluke (Schistosoma mansoni)

RESULT 5
JC5704
protein disulfide-isomerase (EC 5.3.4.1) ER60 precursor - human
N:Alternate names: ER-60 proteinase
C:Species: Homo sapiens (man)
C:Date: 25-Nov-1997 #sequence_revision 25-Nov-1997 #text_change 21-Jan-2000
C:Accession: JC5704; G02888
R:Rutledge, R.; Oda, T.; Ito, H.; Moriyama, T.; Utsumi, S.; Kito, M.
J. Biochem. 122, 634-642, 1997
A:Title: Functions of characteristic Cys-Gly-His-Cys (CGHC) and Gln-Glu-Asp-Leu (QEDL) n
A:Reference number: JC5704; MUID:98060510; PID:9399589
A:Accession: JC5704

C:Keywords: endoplasmic reticulum; intramolecular oxidoreductase; isomerase; redox-act
E1-24/Domain: signal sequence #status predicted <SIG>

F:25-505/Product: protein disulfide isomerase ER60 #status predicted <MA>
F:33-118/Domain: thioredoxin homology <TX1>
F:384-470/Domain: thioredoxin homology <TX2>
F:502-505/Region: endoplasmic reticulum retention signal
F:57-60,406-409/disulfide bonds: redox-active #status predicted

| | | | | |
|-----------------------|------------------|-------------------|------------|-------------|
| Query Match | 4.1%; | Score 169.5; | DB 2; | Length 505; |
| Best Local Similarity | 22.68; | Pred. No. 0.0024; | | |
| Matches 64; | Conservative 59; | Mismatches 127; | Indels 33; | Gaps 10; |

[illegible]

```

RESULT 7
S57942
Protein disulfide-isomerase (EC 5.3.4.1) - Aspergillus niger
C:Species: Aspergillus niger
C:Date: 13-Jan-1996 #sequence_revision 13-Mar-1997 #text_change 16-Jul-1999
C:Accession: S57942
R:Malprichtl, S.
submitted to the EMBL Data Library, July 1995
A:Description: Heterologous expression in Aspergillus.
A:Reference number: S57942
A:Accession: S57942
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-515 <MAL>
A:Cross-references: EMBL:X89449, NID:g889148, PIDs:CA61619.1, PID:g889149
A:Superfamily: Protein disulfide-isomerase; thioredoxin homology
C:Keywords: Intramolecular oxidoreductase; isomerase
/33-116/Domain: thioredoxin homology <TX1>

```

| | | | | |
|-----------------------|------------------|-------------------|-------------|-------------|
| Query Match | 4.1%; | Score 169.5; | DB 2; | Length 515; |
| Best Local Similarity | 20.6%; | Pred. No. 0.0025; | | |
| Matches 111; | Conservative 89; | Mismatches 193; | Indels 145; | Gaps 27; |

```

QY 15 VIMCIPMPVNSJ.PELISPOKVFSTLQGLBELNEAVRPLDQDYSIAKVCNCKEELSR 74
Db 44 LVLAEEFAFPMCGHCKALAP-KY-----EEATTELKANPIPLVAKDCTAAE----- 87
QY 75 CGKEKIDMKAVLEFKNILLREF-----PTDLLEDVNAIVAHVLEALL--FSEVRYTN 125
Db 88 -----DLCSQGVGYPTLKFRCVDSSKPYOGARQTESIVSYMIKSLPAVSVANE-EN 141
QY 126 LEDQNIENMLKGANIIFSVVAIGIPEIRAVMEAGFVYGTYYQVPLTTEIALLSEIGS 185
Db 142 LEEKTMDKTI-----VWIGIIPSDQDETQAEEKVAESORDNLTFAIPDAADAIASSEG 195
QY 186 EDVYVADLY--FFHCKLVLD--LQOCCRTYLMEOPLTTLINILHFIKTMKAPLLEVAEDP 241
Db 196 EQPSIV-LYKPFDEKKAIVYDGEIQEA-----IHSWVASASPIVLGEIG--P 239
QY 242 QQSVTHLQGLPLVLFVSOQATYEADRPAAEW-----VAMRLG--KAGY 285
Db 240 ETVYSG-LVIGAVPLIAYFA--ETKEERKAYEDRKPIAQKHGKAINIATIDAKMFGHAG 297

```

```

OY 286 LLLRLDSLEVINIPD---ANWKEKRAEE--GVFVEEVLVHDVLLISHENNMHHEIQ 339
Db 298 LNLDSQKFPAPFAIDDPKANNAKYPYDQAKELNADVEKEFI--DVLDCGKEPISKSEPV 354
OY 340 EDEDNDMEGPDIDVODDEVAETVFRDRRKRLPELTVETLEETFNATVMASDSIVL--FY 397
Db 355 ESQ-----EGP-----VTV--VVAHSKDLVIDNDKDVLEFY 385
OY 398 AGMQAVNAFLQSYIIVAAVKLKGSTMLLRINCADMSDVCTK-----QNV 443
Db 386 APWCCHCKALAPKDELTALV-----ADHPDLAAKVTAIKATANDVDPPI 432
OY 444 TEPLIKKTKKG--ENPVSYAGMLCTDKLKFLOLNKISTPVNITTSQOEAEVLSGL 499
Db 433 TGFPLLRITPPGAKDSPIEYSGSRFVDDLANFKEEN--KINVALNVAASETOEGGV 489

```

RESULT 8
JC2385
protein disulfide-isomerase (EC 5.3.4.1) ER60 precursor - bovine
NAlternate names: ER60 proteinase
CSpecies: Bos primigenius taurus (cattle)
CDate: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
CAccession: JC2385
RHitano, N.; Shimasaki, F.; Kato, H.; Sakai, R.; Tanaka, T.; Nishida, J.; Yazaki, Y.
Biochem. Biophys. Res. Commun. 204, 375-382, 1994
ATitle: Molecular cloning and characterization of a cDNA for bovine phospholipase C-
Reference number: JC2385; MUID:95032122; PMID:7945384

A:Residues: 1-488 <HR>
A:Cross-references: DDBJ:D16234
A:Experimental source: thymus
C:Comment: ER60 may be part of a complex capable of catalyzing cleavage of itself, of
C:Comment: This enzyme has two Trp-Cys-Gly-His-Cys-Iys motifs.
C:Superfamily: protein disulfide-isomerase; thioredoxin homology
C:Keywords: duplication; endoplasmic reticulum; intramolecular oxidoreductase; isomer
F:1-24/Domain: signal sequence #status predicted <IG>
F:25-488/Product: ER60 protease #status predicted <MAT>
F:33-118/Domain: thioredoxin homology <TX1>
F:367-453/Domain: thioredoxin homology <TX2>
F:485-488/Region: endoplasmic reticulum retention signal
F:57-60,389-392/Disulfide bonds: redox-active #status predicted

Query Match 4.0%; Score 167; DB 1; Length 488;
Best Local Similarity 23.0%; Pred. No. 0.0033;
Matches 64; Conservative 53; Mismatches 121; Indels 40; Gaps 9;

| | | | |
|---|--------|--|-----|
| Oy | 376 | VEEFEENFNAIVMASSD-----IYLFVGMQAVASNAFLQSIVDAVAVKIKGSTMILRINC | 431 |
| Dd | 28 | LHLDIDNNESKRIITDTSGSLMLVEFFAPWGHCKKLAPETEAATRLKG--IYLPLAKVD | 85 |
| Oy | 432 | ADMSDVCTKWVFEEPIIKMKKKENDSVSAGMGITDLKLTQLNRISYPVNITSIOEA | 491 |
| Dd | 86 | TANINNTCKKVGVCSGYPLTLIRPDEEGSAVGDPRTAIGVSHLKQAQPASVPFKSEEF | 145 |
| Oy | 492 | EELTSELGLKPILLIYSVSVLGFSPIMKTAKEDFSAGNYLKGVIYTGISSEDDULLS | 551 |
| Dd | 146 | EKFISDK-----DASYVGFEPKDLFESAHSFTLAASNLRD---NKRFAHTNWESLV | 193 |
| Oy | 552 | TKYAASIPALLAR--HTEGKIIESIPLASTHAODIVOILTALLEMPETITVENLSYPR | 609 |
| Dd | 194 | NKYDDDGEGITLFRPSHLTNKEFKTYAYIE-----QKMISGIKRFIQ---EN----- | 239 |
| Oy | 610 | LQRPLLILESD----GTVNPQQYRKALITLVOKRYIDS | 642 |
| Dd | 240 | GKDILLAYVDVDEKNAGSNMYRNRMVMAMAKFLDA | 276 |
| RESULT | 9 | | |
| protein disulfide-isomerase (EC 5.3.4.1) ER60 precursor - human | 555307 | | |

Arch. Biochem. Biophys. 308, 454-460, 1994
A:Title: Ekp61 is Grp58, a stress-inducible luminal endoplasmic reticulum protein, but 1
A:Reference number: S41661; MUID:94153092; PMID:8109975
A:Accession: S41661
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-504 <MA>
C:Superfamily: protein disulfide-isomerase; thioredoxin homology
C:Keywords: endoplasmic reticulum; intramolecular oxidoreductase; isomerase; redox-acti
F:33-118/Domain: thioredoxin homology <TX1>
F:383-469/Domain: thioredoxin homology <TX2>

Query Match 3.7%; Score 152; DB 2; Length 504;
Best Local Similarity 20.7%; Pred. No. 0.034;
Matches 61; Conservative 53; Mismatches 123; Indels 58; Gaps 9;

Qy 376 VELTETENATYMASDS----VLFYAGQAVSNAPFLOSIDYAVKLGKSTMLTRINC 431
Db 28 LETDENESRVSDFGAGMLVEFFAPWCCHKRLAPEYEAATRLKG--TVPLAKYDC 85
Qy 432 ADMSPVCTKONVTEPPIIKMYKKGPNVSYAGMLGKTKDLKFIOLNRISYPVNTSIOEA 491
Db 86 TANTMTCKKYGTGYPTLIKIFPDGEAGAYDGPRTADGIVS--HLKKQAGPASVPLRTED 143
Qy 492 EYLSGELYKDLILYSSVVLGLFSPMTKTAKEDSEAGNYLKQYV-----ITGIYSEE 545
Db 144 E-----FKKFISDKDASVGVGFPRDLFSDGSEFLKAASNLRDVYRFAHTVWESLVKRY 196
Qy 546 D-----VLLSTYVYASLPLALLARHTGKTEST-----PLASTHAODIYOI 587
Db 197 DONGGIRIFRPLHLANKFEEDKIVAYTEKMTSGKSRSFLFRKAFGLCPHNTEDNKDLIO- 255
Qy 588 ITDALLEMPETVNLPSYFRLKPLILFSDGTVPNQYKAILTVYKOKLDS 642
Db 256 -----GKDLTAIYDVD-----YKNTKGSNMYNRVMMVAKTFLDA 292

RESULT 14

JC7623
protein disulfide-isomerase (EC 5.3.4.1) precursor - yeast (Pichia pastoris)
C:Species: Pichia pastoris
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 24-Aug-2001
C:Accession: JC7623
R:Masame, A.; Vad, R.; Kristensen, T.; Oyen, T.B.
Biochem. Biophys. Res. Commun. 281, 1176-1182, 2001
A:Title: Characterization of a gene encoding a Pichia pastoris protein disulfide isomera
A:Reference number: JC7623; MUID:21139750; PMID:11243858
A:Accession: JC7623
A:Molecule type: DNA
A:Residues: 1-517 <MA>
A:Cross-references: GB:AJ302014
A:Experimental source: strain GS115
C:Genetics:
A:Gene: pdt
C:Superfamily: protein disulfide-isomerase; thioredoxin homology
C:Keywords: intramolecular oxidoreductase; isomerase; multifunctional enzyme
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-517/Product: protein disulfide-isomerase #status predicted <MAT>
F:59-66/Domain: redox-active #status predicted <REAI>
F:401-408/Domain: redox-active #status predicted <REAI>
F:514-517/Region: endoplasmic reticulum retrieval signal #status predicted

Query Match 3.6%; Score 150.5; DB 2; Length 517;
Best Local Similarity 22.5%; Pred. No. 0.044;
Matches 79; Conservative 64; Mismatches 161; Indels 47; Gaps 15;

Qy 376 VELTETENATYMASDSIVL--FYAGQAVSNAPFLOSIDYAVKLGKSTMLTRINCADW 434
Db 36 VLTETAFESFTSNPHVLAEPFAPWCCHKKLGPELVSAAILDNDEQVIAQIDCTEE 95
Qy 435 SDVCTKONVTEPPIIKMYK--KGENPVSYAGMLGKTKDLKFIOLNRISYPVNTSIOEAE 493
Db 96 KELCGYEIKGYPTLVKFGVEVPSDYQGROGOSIYSYMLKQSLPVPSEINATKDLDD 155

Qy 494 YLSGELYKDLILYSSVVLGLFSPMTKTAKEDSEAGNYLKQYVITGIYSEE--DVLLIS 551
Db 156 TIA-EAKPEVI-----YQVL-----PEASNLESNTTGYAGATLREKFTFSTMS 200
Qy 552 TRYAA-----SLPALLARHTGKTEST-----PLASTHAODIYOITDALLEMPETIVE 602
Db 201 TVYAKKRYTSDSTNPAYLVLR--PGEEPSVYSGEELDETHLVHWIDIESKPL---FGDIDGS 255
Qy 603 NLPYFRLQKPLILFSDGTVPNQYKAILTVK-----QKLDSTFPCWMLNKTPVGR 657
Db 256 TFKSYAEANIPLAYYYFE--NEEQAAADAIKPKAKQKRIKF---VGLDAVKFK 308
Qy 658 GILRAYFDPLPLVLIVNLHSGGVFAFPDQAIIEENLVMLKLEAG 708
Db 309 HAKNLNMBE-EKLPLEVHIDLVS--NKKFGVPQDQELTNKDVLELLEKFIAG 357

RESULT 15

T06724
protein disulfide-isomerase homolog F28P10.60 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Oct-1999
C:Accession: T06724
R:Queller, F.; Choisme, N.; Robert, C.; Brotlier, P.; Wincker, P.; Catolico, L.; Art
submitted to the Protein Sequence Database, April 1999
A:Reference number: Z15793
A:Accession: T06724
A:Molecule type: DNA
A:Residues: 1-566 <QE>
A:Cross-references: EMBL:AL049655; GSPDB:GN00061; ATSP:F28P10.60
A:Experimental source: cultivar Columbia; BAC clone F28P10
C:Genetics:
A:Gene: ATSP:F28P10.60
A:Map position: 3
A:Introns: 192/2; 235/3; 291/1; 330/3; 351/3; 380/3; 405/3; 428/3; 455/3; 493/3; 512/
C:Superfamily: protein disulfide-isomerase; thioredoxin homology

Query Match 3.6%; Score 148; DB 2; Length 566;
Best Local Similarity 20.5%; Pred. No. 0.074;
Matches 129; Conservative 71; Mismatches 200; Indels 228; Gaps 28;

Qy 286 LILLRDSLEVNIPQ-----DANVVFKAEEGVVEFLVLDVLIISHVNNHIEIORD 341
Db 17 ILLVNSRRENMSSGSDDELAFLAAESKQ-----SHGGSYH-EEENH 63
Qy 342 EDNDMEG-----PDIVDDDEVAETVRRRRKPLLELVEL 378
Db 64 QHRDEFENYDLEQGGGEFHNGDHYEEBPLPVDEKDVAV-----L 104
Qy 379 TREETENATY--MASDSIVL--FYAGQAVSNAPFLOSIDYAVKLGKSTMLTRINCADMSV 437
Db 105 TKDNFTFVGNNSFAMVEFYAPWCACQALTPETVYAAATLKGALAA--LAKIDATEEGDL 162
Qy 438 CTKONVTEPPIIKMYKKGPNVSYAGMLGKTKDLKFIOLNRISYPVNTSIOEAEVYLSG 497
Db 163 ACKYELGPGPYFLFVVDGEMRKTYEGEERKDDIVMLKKKASPISHNITTKREAEVLSA 222
Qy 498 E-----LYKDLILYSSVS--VLGLF-----PYMKTAKE 524
Db 223 EKLVLGFLNLSVSGSEELAASRLDLSFYOTASPDIARLFETETQVKKRPALVLK 282
Qy 525 DFESE-----GNVLKGIVITGIYSEEDVILLS--TKYAA-----LPALL 562
Db 283 EEEKLAREDNFTKTAIAEFVSAKVPVIVNTRBQASLIFSSVKNQANESKHLPTLR 342
Qy 563 -LARTTEKRI-----ESIP-----LASTHAODIYOITDALLEMF 596
Db 343 EVAKSPKGFVYVYQMONEDYGEAVSGFVGTAAPKLVLYTGNEDMKFLLDG----- 397
Qy 597 PEITVENLPSYFRLQKPLILFSDGTVPNQY-----KAI-----LTVKQ 637
Db 398 -ELTVNNI-----KTLAEFLADKLKPFYKSDPLPENNDGVKIVGNNEDEIVLDES 449

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:10:57 ; Search time 16 Seconds
(without alignments)
2091.962 Million cell updates/sec

Title: US-09-847-046-2
Perfect score: 4142

Sequence: 1 MFSGFNFRVIGISFVIMCIF.....FKEAKSPRRDKELGCSKVN 807

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SWISSPROT_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|------------|---------------------|
| 1 | 185 | 4.5 | 505 | PDI_HUMAN | P5059 humicola ln |
| 2 | 177 | 4.3 | 496 | PDI_DROME | P54399 drosophila |
| 3 | 175.5 | 4.2 | 505 | PDA3_BOVIN | P38657 bos taurus |
| 4 | 172.5 | 4.2 | 515 | PDI_ASPO | Q00248 aspergillus |
| 5 | 172.5 | 4.2 | 618 | PDA4_CAEL | P34329 caenorhabdi |
| 6 | 169.5 | 4.1 | 505 | PDA3_HUMAN | P30101 homo sapien |
| 7 | 169.5 | 4.1 | 515 | PDI_ASPO | P11598 rattus norv |
| 8 | 163.5 | 3.9 | 504 | PDA3_MOUSE | P27773 mus musculu |
| 9 | 162 | 3.7 | 522 | PDI_YEAS | P1967 saccharomyc |
| 10 | 154.5 | 3.5 | 497 | BS2_TRYB | P12865 trypanosoma |
| 11 | 147 | 3.5 | 492 | PDI1_SCHPO | Q10057 schizosacch |
| 12 | 146.5 | 3.5 | 492 | PDI1_SCHPO | P13667 homo sapien |
| 13 | 142 | 3.4 | 643 | PDA4_RAT | P36659 rattus norv |
| 14 | 138.5 | 3.3 | 645 | PDA4_HUMAN | P38658 schistosoma |
| 15 | 138 | 3.3 | 484 | ER60_SCHMA | P38920 saccharomyc |
| 16 | 137 | 3.3 | 769 | MEH1_YEAS | P38194 saccharomyc |
| 17 | 137 | 3.3 | 2493 | PDA4_MOUSE | P22623 arabidopsis |
| 18 | 136 | 3.3 | 509 | PDI_RABIT | Q00000 arabidopsis |
| 19 | 135.5 | 3.3 | 509 | PDI_RAT | P21195 oryctolagus |
| 20 | 135 | 3.3 | 508 | PDI_HUMAN | P04785 rattus norv |
| 21 | 135 | 3.3 | 510 | PDI_BOVIN | P07237 homo sapien |
| 22 | 134.5 | 3.2 | 517 | EU61_YEAS | P34474 saccharomyc |
| 23 | 133.5 | 3.2 | 638 | PDA4_MOUSE | P08003 mus musculu |
| 24 | 133 | 3.2 | 361 | PDA6_ARATH | O22263 arabidopsis |
| 25 | 132 | 3.2 | 2869 | RBP1_PLAIV | Q00000 arabidopsis |
| 26 | 131.5 | 3.2 | 364 | PDA6_MEDSA | P38661 medicago sa |
| 27 | 131 | 3.2 | 1088 | RPO_ROT51 | P22678 simian 11 r |
| 28 | 127.5 | 3.1 | 1905 | Y659_PASMU | Q96m21 pasteurella |
| 29 | 127 | 3.1 | 2108 | NOT1_YEAS | P25655 saccharomyc |
| 30 | 126 | 3.0 | 493 | PDI_CHICK | P03102 gallus galli |
| 31 | 125 | 3.0 | 930 | DPO1_HAETN | P43741 haemophilus |
| 32 | 124.5 | 3.0 | 509 | PDI_MOUSE | P09103 mus musculu |
| 33 | 123.5 | 3.0 | 501 | PDI1_ARATH | O9x101 arabidopsis |

| | | | | | | |
|----|-------|-----|------|---|------------|--------------------|
| 34 | 123 | 3.0 | 644 | 1 | C3A_BACTT | P07130 bacillus th |
| 35 | 122.5 | 3.0 | 4451 | 1 | GRSB_BACBR | P14688 b gramici |
| 36 | 122 | 2.9 | 586 | 1 | UVRC_MYCPN | P75350 mycoplasma |
| 37 | 121.5 | 2.9 | 1088 | 1 | RPO_ROTBR | P17468 bovine rota |
| 38 | 121.5 | 2.9 | 3911 | 1 | AKA9_HUMAN | Q09996 h-a-kinase |
| 39 | 120 | 2.9 | 1102 | 1 | UBE7_HUMAN | Q03009 homo sapien |
| 40 | 119.5 | 2.9 | 359 | 1 | PDI2_SCHPO | Q13811 schizosacch |
| 41 | 118.5 | 2.9 | 519 | 1 | PDA5_HUMAN | Q14554 homo sapien |
| 42 | 118.5 | 2.9 | 1014 | 1 | NEB1_HUMAN | Q76041 homo sapien |
| 43 | 118.5 | 2.9 | 1088 | 1 | RPO_ROTBU | P21615 bovine rota |
| 44 | 116 | 2.8 | 341 | 1 | MO25_HUMAN | Q09376 homo sapien |
| 45 | 116 | 2.8 | 1818 | 1 | Z294_HUMAN | O94822 homo sapien |

ALIGNMENTS

| RESULT 1 | ID | PDI_HUMAN | STANDARD: | PRT: | 505 AA. |
|----------|--|-----------|-----------|------|--|
| AC | P55059: | | | | |
| DT | 01-OCT-1996 (Rel. 34, Created) | | | | |
| DT | 01-OCT-1996 (Rel. 34, Last sequence update) | | | | |
| DT | 01-OCT-1996 (Rel. 34, Last annotation update) | | | | |
| DE | Protein disulfide isomerase precursor (Ec 5.3.4.1) (PDI). | | | | |
| OS | Humicola, Fungi; Ascomycota; mitosporic Ascomycota; Humicola. | | | | |
| OX | NCBI_TaxID=34413; | | | | |
| RM | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN-KASI: | | | | |
| RX | MEDLINE=94369094; PubMed=7765273; | | | | |
| RA | Kajino T., Sarai K., Imada T., Idekoba C., Asami O., Yamada Y., | | | | |
| RA | Hirai M., Ueda S.; | | | | |
| RT | "Molecular cloning of a fungal cDNA encoding protein disulfide | | | | |
| RT | isomerase."; | | | | |
| RT | Biosci. Biotechnol. Biochem. 58:1424-1429(1994). | | | | |
| CC | -I- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING | | | | |
| CC | DISULFIDE BONDS. MAY BE INVOLVED IN GLYCOSYLATION. PROLYL | | | | |
| CC | HYDROXYLATION AND TRIGLYCERIDE TRANSFER (BY SIMILARITY). | | | | |
| CC | -I- CATALYTIC ACTIVITY: Rearrangement of both intrachain and | | | | |
| CC | interchain disulfide bonds in proteins to form the native | | | | |
| CC | structures. | | | | |
| CC | -I- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (By similarity). | | | | |
| CC | -I- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS. | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | | |
| CC | or send an email to license@isb-sib.ch). | | | | |
| DR | EMBL: S74296; AAC60578.1; - | | | | |
| DR | HSSP: P07237; IMEX. | | | | |
| DR | InterPro: IPR000886; ER_target. | | | | |
| DR | InterPro: IPR000063; ThioRed. | | | | |
| DR | Pfam: PF00085; ThioRed; 2. | | | | |
| DR | PRINTS: PR00421; THIOREDOXIN. | | | | |
| DR | TIGRFAMS: TIGR01126; pdi_dom; 2. | | | | |
| DR | TIGRFAMS: TIGR01130; ER_pdi_fam; 1. | | | | |
| DR | PROSITE: PS00014; ER_TARGET; 1. | | | | |
| DR | PROSITE: PS00194; THIOREDOXIN; 2. | | | | |
| KW | Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal. | | | | |
| FT | SIGNAL | 1 | 20 | | POTENTIAL. |
| FT | CHAIN | 21 | 505 | | PROTEIN DISULFIDE ISOMERASE. |
| FT | DISULFID | 50 | 53 | | REDOX-ACTIVE (BY SIMILARITY). |
| FT | DISULFID | 385 | 388 | | REDOX-ACTIVE (BY SIMILARITY). |
| FT | SITE | 502 | 505 | | PREVENT SECRETION FROM ER (POTENTIAL). |
| SQ | SEQUENCE | 505 AA; | 55114 MW; | | 2767CODI6IB5F95 CRC64; |

Query Match

4.5%; Score 185; DB 1; Length 505;

| | | | | | | | | | | | | | |
|------------|---|--|------|------------|------|--------|------|------|-----|--|--|--|--|
| Matches | 99: | Conservative | 96: | Mismatches | 180: | Indels | 122: | Gaps | 25: | | | | |
| QY | 373 | ETLVELTEETENAVY-----MASDSIVL--FYAGQWAVSAFLQSYIDVAVKL-KGTST | 423 | | | | | | | | | | |
| Db | 20 | EAEVVEEGVLAVADVNFNEKQLIADNDFVLEVEYAPWCGHCKALAPRYAKAQAQLKEKSP | 79 | | | | | | | | | | |
| QY | 424 | MULTITINCADNSDVCTKONVTEFFPIITKMKKEENVSYTAGHGTDLKFLIDNASTIPV | 483 | | | | | | | | | | |
| Db | 80 | IKLAVVDITVEGELAEQYAVRGVPPILKEFRSG-SPVEYSGGROADIDIAWYTKTKTGPAPK | 138 | | | | | | | | | | |
| QY | 484 | NITTSIOEAEEVYSGELAYKDLILYSVSVYGLFSPMTKAKEDFESAGNYLKGVTGTAYS | 543 | | | | | | | | | | |
| Db | 139 | DITSDADAEQFL-----KD-----NETAIGFKKDLSEAPAKFTFTVANAALDSFV-GYSS | 188 | | | | | | | | | | |
| QY | 544 | EEDVILLSTRKYVASLPALLARHTTEGKTESIPLASTHADIVQIITDALLMEFEITVEN | 603 | | | | | | | | | | |
| Db | 189 | NADVY--AKYEAKNQGVLFKFPFDK-----KSVEF--GELNEB | 224 | | | | | | | | | | |
| QY | 604 | LPSTYRLQK-PLILFSDGTVAPQY-----KKAILLVYQKYLIDSTPPCMNLKNTPVGRG | 658 | | | | | | | | | | |
| Db | 225 | LKKFQAVOSLPILTVDFNHESASKIGGSIKSHLFEVSRE-----GG | 266 | | | | | | | | | | |
| QY | 659 | ILRAEFDLPPLPL-----LVLVNLHSGQ-----YFAF----- | 687 | | | | | | | | | | |
| Db | 267 | HEIKVYDPLKELAKKYYRDIILFTVITSSDEEDHTRIFEEFGMKKEVPTRILKLEEDMAK | 326 | | | | | | | | | | |
| QY | 688 | --PSQDAIIEENVLMULK-LEAGLENNHTT--LPAQEW-KRP--LPAVDFLSM-IDAA | 737 | | | | | | | | | | |
| Db | 327 | YKPESDDIASAETIEAFILKKFLDKGLKHLLSQELP-EDWDKNPVKVLVSSNFEESVALDKS | 385 | | | | | | | | | | |
| QY | 738 | TSQRTKRPV---KMKELDVQENKEQEHDEKSAV-----RKPEIPLRKHHNRSMW | 787 | | | | | | | | | | |
| Db | 386 | KSVLVEFYAPWCGHCKQLAPYIDQLAEYKQDNEDIVIAKMDSTANLESLIKISSPTIKY | 445 | | | | | | | | | | |
| QY | 788 | FKEAEK---SFRDKEL_801 | | | | | | | | | | | |
| Db | 446 | PKEDNKVIDENLDRTL_462 | | | | | | | | | | | |
| RESULT 3 | | | | | | | | | | | | | |
| PDA3_BOVIN | | | | | | | | | | | | | |
| ID | PDA3_BOVIN | STANDARD: | PRT: | 505 | AA. | | | | | | | | |
| AC | P38657: | | | | | | | | | | | | |
| DT | 01-FEB-1995 (Rel. 31, Created) | | | | | | | | | | | | |
| DT | 01-FEB-1995 (Rel. 31, Last sequence update) | | | | | | | | | | | | |
| DT | 16-OCT-2001 (Rel. 40, Last annotation update) | | | | | | | | | | | | |
| DE | Protein disulfide isomerase A3 precursor (EC 5.3.4.1) (Disulfide | | | | | | | | | | | | |
| DE | isomerase ER-60) (ERP60) (58 kDa microsomal protein) (P58) (ERP57). | | | | | | | | | | | | |
| CN | PDA3 OR GRP58. | | | | | | | | | | | | |
| OS | Bos taurus (Bovine). | | | | | | | | | | | | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | | | | | | | | | | | |
| OC | Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; | | | | | | | | | | | | |
| OC | Bovidae; Bovinae; Bos. | | | | | | | | | | | | |
| NCBI_OX | NCBI_TaxID=9913; | | | | | | | | | | | | |
| NCBI_OX | [1] | | | | | | | | | | | | |
| RP | SEQUENCE FROM N.A. | | | | | | | | | | | | |
| RA | MEDLINE=96096758; PubMed=8529662; | | | | | | | | | | | | |
| RA | Hirano N., Shibasaki F., Sakai R., Tanaka T., Nishida J., | | | | | | | | | | | | |
| RA | Yazaki Y., Takenawa T., Hirai H.; | | | | | | | | | | | | |
| RT | "Molecular cloning of the human glucose-regulated protein | | | | | | | | | | | | |
| RT | ERP57/GRP58, a thiol-dependent reductase. Identification of its | | | | | | | | | | | | |
| RT | secretory form and inducible expression by the oncogenic | | | | | | | | | | | | |
| RT | transformation."; | | | | | | | | | | | | |
| RL | Eur. J. Biochem. 234:336-342(1995). | | | | | | | | | | | | |

| | |
|--------------------------|--|
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation |
| CC | the European Bioinformatics Institute. There are no restrictions on its |
| CC | use by non-profit institutions as long as its content is in no way |
| CC | modified and this statement is not removed. Usage by and for commercial |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ |
| CC | or send an email to license@isb-sib.ch). |
| CC | ----- |
| CC | EMBL; D16235; BAA03760.1; -. |
| DR | HSSP; P07237; IMEX. |
| DR | InterPro; IPR000063; ThioRed. |
| DR | Pfam; PF00085; ThioRed; 2. |
| DR | PRINTS; PR00421; THIOREDOXIN. |
| DR | TIGRFAMS; TIGR01126; pd1_dom; 2. |
| DR | TIGRFAMS; TIGR01130; ER_pdl_fam; 1. |
| DR | PROSITE; PS00194; THIOREDOXIN; 2. |
| KW | Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal. |
| FT | SIGNAL 1 24 BY SIMILARITY. |
| FT | CHAIN 25 505 PROTEIN DISULFIDE ISOMERASE A3. |
| FT | DISULFD 57 60 REDOX-ACTIVE (BY SIMILARITY). |
| FT | DISULFD 406 409 REDOX-ACTIVE (BY SIMILARITY). |
| FT | Site 502 505 PREVENT SECRETION FROM ER (BY SIMILARITY). |
| FT | Site 502 505 |
| SQ | SEQUENCE 505 AA; 56929 MW; A85911748DC70A23 CRC64; |
| Query Match | 4.2%; SCORE 175.5; DB: 1; Length 505; |
| Best Local Similarity | 23.3%; Pred. No. 0.00037; |
| Matches 66; Conservative | 56; Mismatches 128; Indels 33; Gaps 9; |
| OY | 376 VELTEEFENATVMASSD---IVLFFYAGQAVSMALQSYDVAVKLTGSTMILTRINC 431 |
| DB | 28 LEITDNEESRTDTPGGSLMVEFAWCGCKRLAEYEAAARLKG--IVPLAKDC 85 |
| OY | 432 ADMVDVCTKQNVTEPPIIKMKKGNPNVSYAGMLGTOKDLFLFIQLNRISYPNITSIGA 491 |
| DB | 86 TANTNTCKNKYSGVPTLKIFRDGEESGAYDGPRTADGIVSHLKKQAGPASVPLKSEEF 145 |
| OY | 432 EETLSGELTKDLLKSYSVGLFSEPTMKTKADEPSEAGNIYKGVITGIYSEEDVLLS 551 |
| DB | 146 EKFIIDK-----DASVGFEDLDFSEAHSEFLKAASNRD--NYRFAHTVESLV 193 |
| OY | 552 TVYASLPALLLAR--HREGKTESIPLASTHQ----DIVQITDALLEMPETIVENLP 605 |
| DB | 194 NKYDDGEGITLFRSHLTNKFEDKTVAYTEKMTSGIKRRIQINIFIGICPHMEDKND 253 |
| OY | 606 SYFRLQ-KPLLIFSD---GTVNPQYKKAILLVKQKYIDS 642 |
| DB | 254 ---LLQKDLITAYVDVDEKNAKGSNYWRNVMVAKKFLDA 293 |
| RESULT 4 | |
| PDI_ASPOR | |
| ID | PDI_ASPOR STANDARD; PRT: 515 AA. |
| AC | 000248; |
| DT | 01-NOV-1997 (Rel. 35, Created) |
| DT | 01-NOV-1997 (Rel. 35, Last sequence update) |
| DT | 30-MAY-2000 (Rel. 39, Last annotation update) |
| DE | Protein disulfide isomerase precursor (EC 5.3.4.1) (PDI). |
| GN | PDI. |
| OS | Aspergillus oryzae. |
| OC | Eukaryota; Fungi; Ascomycota; Pezilomycolina; Eurotiomycetes; |
| OC | Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus. |
| OX | NCBI_TaxID=5062; |
| RN | (1) |
| RP | SEQUENCE FROM N.A. |
| RC | STRAIN-RIB 40; |
| RA | Lee B., Yamada O., Kitamoto K., Takahashi K.; |
| RL | Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases. |
| CC | -I- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING |
| CC | DISULFIDE BONDS, MAY BE INVOLVED IN GLYCOSTYLATION, PROLYL |
| CC | HYDROXYLATION AND TRIGLYCERIDE TRANSFER (BY SIMILARITY). |
| CC | -I- CATALYTIC ACTIVITY: Rearrangement of both intrachain and |
| CC | interchain disulfide bonds in proteins to form the native |
| CC | structures. |

Db 114 DKGPNVDGGRDAGIVENWESRYDPNKKPPEEVVLTLTENNEDPDISNNELVLEFYA 173
 QY 399 GMAVAMAFLOSIDYIVAVKIKGT-STMLTTRINCADMSQVCKQVTERPIIKMKKKGN 457
 Db 174 PWCCKLAPLEYEKAQKAKQSKVKLGKVDATIEKDLGKGVSGIPMKIIRNGR 233
 QY 458 PVSVAAGMLGTDLKFLIOLNRIISYPVNTSIOEAEYLSGELIKDLIIYSVSVGLFSP 517
 Db 224 -FDVYGPFEAAGIIKYMIDQSKPAKKLPKLKDVRFMS-----KD-----DVTIIGFAT 283
 QY 518 TMTAKKEPSEAGNYLKGVIITGISEE-----DVLILSTKIASPLLLAR 565
 Db 284 EDSTAFEAFSDAEMLRREEFKTMGHTSDPAFKKWDKPNDIITIE-----YPSLFRSK 336
 QY 566 HTEGKIESIPLASTFAODIYOITLTDALTEMPPEITVENLPSYFRLOKPLLIIFSGTVP 625
 Db 337 F-EPKSRITNNKAATSEDLAFEFREHSAPLYGKMTKKNATRY-TKKPLVYVYVYNDSSV 394
 QY 626 QYKKA-----ILTLYKQKYLDSFTPCWLNKNTPVGRGILRAYFDPLPLLVNL 678
 Db 395 QYREGSEYWRSKVLINIAQKQKDKYK----- 420
 QY 679 HSGGVAFEPDQATIEENLVMLKL---EAGLENHTITLPAQEMKPPPLPYDLSMID 735
 Db 421 -----FAVADEEFRAKE-----LEELIGDSGLHNHYVFGYDGKKYPMNDEDFEID 469
 QY 736 ---AATSGRGTRKVRKCMKERTVOENDK 760
 Db 470 ENLEAFMKQISSGKAKAHVKSAPKADK 498

RESULT 6

PD03_HUMAN STANDARD: PRT: 505 AA.
 ID PD03_HUMAN Q13453; Q99UM7;

AC P30101; Q14255; Q13453; Q99UM7;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein disulfide isomerase A3 precursor (EC 5.3.4.1) (Disulfide isomerase ER-60) (ERP60) (58 kDa microsomal protein) (p58) (Erp57)
 DE (58 kDa glucose regulated protein).
 GN PDIA3 OR GRP58 OR ERP60.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95032122; PubMed=7945384;
 RA Hirano N., Shibasaki F., Katoh H., Sakai R., Tanaka T., Nishida J., Yasaki Y., Takenawa T., Hirai H.;
 RT "Molecular cloning and characterization of a cDNA for bovine phospholipase C-alpha: proposal of redesignation of phospholipase C-alpha.";
 RT Biochem. Biophys. Res. Commun. 204:375-382(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Charneck-Jones D.;
 RT Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Kito M., Urae R.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=96063616; PubMed=7487104;
 RA Bourd M., Demady D., Martin J.L., Jabbour S.K., Martin B.M., George J.W., Pohl U.R.;
 RT "cDNA cloning and baculovirus expression of the human liver endoplasmic reticulum p58: characterization as a protein disulfide isomerase isoform, but not as a protease or a carnitine

RT acyltransferase.";
 RL Arch. Biochem. Biophys. 323:397-403(1995).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97349107; PubMed=9205111;
 RA Kojunen P., Horelli-Kuitunen N., Heiaakoski T., Karvonen P., Jaakkola M., Palotie A., Kivirikko K.I.;
 RT "Structures of the human gene for the protein disulfide isomerase-related polypeptide ERp60 and a processed gene and assignment of these genes to 15q15 and 1q21.";
 RT Genomics 42:397-404(1997).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RA Strausberg R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 25-33.
 RC TISSUE=Liver;
 RX MEDLINE=93162045; PubMed=1286669;
 RA Hochstrasser D.F., Frutiger S., Paquet N., Bairoch A., Ravier F., Pasquali C., Sanchez J.-C., Tissot J.-D., Bjellqvist B., Vargias R., Appel R.D., Hughes G.J.;
 RT "Human liver protein map: a reference database established by microsequencing and gel comparison.";
 RT Electrophoresis 13:992-1001(1992).
 RN [8]
 RP SEQUENCE OF 26-42.
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=97295304; PubMed=9150946;
 RA Rasmussen R.K., Ji H., Edes J.S., Moritz R.L., Reid G.E., Simpson R.J., Dorow D.S.;
 RT "Two-dimensional electrophoretic analysis of human breast carcinoma proteins: mapping of proteins that bind to the SH3 domain of mixed lineage kinase MLK2.";
 RT Electrophoresis 18:588-598(1997).
 RN [9]
 RP SEQUENCE OF 25-54; 62-75 AND 95-104.
 RC TISSUE=Colon carcinoma;
 RX MEDLINE=97295306; PubMed=9150948;
 RA Ji H., Reid G.E., Moritz R.L., Edes J.S., Burgess A.W., Simpson R.J.;
 RT "A two-dimensional gel database of human colon carcinoma proteins.";
 RT Electrophoresis 18:605-613(1997).
 RN [10]
 RP SEQUENCE OF 95-104 AND 472-479.
 RC TISSUE=Keratinocytes;
 RX MEDLINE=93162043; PubMed=1286667;
 RA Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E., Vandekerckhove J.;
 RT "Microsequences of 145 proteins recorded in the two-dimensional gel protein database of normal human epidermal keratinocytes.";
 RT Electrophoresis 13:960-969(1992).
 RN [11]
 RP MASS SPECTROMETRY.
 RC TISSUE=Breast cancer;
 RX MEDLINE=21829512; PubMed=11840567;
 RA Harris R.A., Yang A., Stein R.C., Lucy K., Brusten L., Herath A., Parekh R., Waterfield M.D., O'Hare M.J., Neville M.A., Page M.D., Zvelebil M.J.;
 RT "Cluster analysis of an extensive human breast cancer cell line protein expression map database.";
 RT Proteomics 2:212-223(2002).
 RN [12]
 RP CATALYTIC ACTIVITY: Rearrangement of both intrachain and interchain disulfide bonds in proteins to form the native structures.
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (By similarity).
 CC -1- MASS SPECTROMETRY: MW=54265.22; METHOD=MALDI.
 CC -1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
 CC -1- CAUTION: WAS ORIGINALLY THOUGHT TO BE A PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE TYPE I (PHOSPHOLIPASE C-ALPHA).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: D16234; BAA03759.1; -
 DR EMBL: 249835; CAA08996.1; -
 DR EMBL: D83485; BAA11928.1; -
 DR EMBL: U42068; AAC50331.1; -
 DR EMBL: U75885; AAC51518.1; -
 DR EMBL: U75875; AAC51518.1; JOINED.
 DR EMBL: U75876; AAC51518.1; JOINED.
 DR EMBL: U75877; AAC51518.1; JOINED.
 DR EMBL: U75878; AAC51518.1; JOINED.
 DR EMBL: U75879; AAC51518.1; JOINED.
 DR EMBL: U75880; AAC51518.1; JOINED.
 DR EMBL: U75881; AAC51518.1; JOINED.
 DR EMBL: U75882; AAC51518.1; JOINED.
 DR EMBL: U75883; AAC51518.1; JOINED.
 DR EMBL: BC014433; AAC51518.1; JOINED.
 DR HSSP: P07237; IMEX.
 DR MEROPS: C17.001; -
 DR SWISS-2DPAGE: P30101; HUMAN.
 DR Aairhus/Ghent-2DPAGE; 5410; IEF.
 DR PHCI-2DPAGE; P30101; -
 DR Slena-2DPAGE; P30101; -
 DR Genew; HGNC:4606; GRP58.
 DR MIM: 602046; -
 DR InterPro: IPR000063; ThioRed.
 DR Pfam: PF00085; ThioRed; 2.
 DR PRINTS: PRO0421; THIOREDOXIN.
 DR TIGRFAWS: TIGR01126; pdi_dom; 2.
 DR TIGRFAWS: TIGR01130; ER_PDI_fam; 1.
 DR PROSITE: PS00194; THIOREDOXIN; 2.
 DR Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.
 KW CHAIN 1 24
 FT DISULFID 25 505 PROTEIN DISULFIDE ISOMERASE A3.
 FT DISULFID 57 60 REDOX-ACTIVE (BY SIMILARITY).
 FT SITE 406 409 REDOX-ACTIVE (BY SIMILARITY).
 FT SITE 502 505 PREVENT SECRETION FROM ER
 (BY SIMILARITY).
 FT CONFLICT 19 19 A -> G (IN REF. 3).
 FT CONFLICT 22 22 A -> V (IN REF. 3).
 FT CONFLICT 217 217 D -> Y (IN REF. 1).
 FT CONFLICT 225 225 O -> P (IN REF. 2).
 FT CONFLICT 238 238 E -> G (IN REF. 2).
 FT CONFLICT 272 272 N -> D (IN REF. 1 AND 5).
 FT CONFLICT 355 355 D -> G (IN REF. 1).
 FT CONFLICT 358 358 D -> G (IN REF. 1).
 FT CONFLICT 368 368 E -> D (IN REF. 1).
 SQ SEQUENCE 505 AA; 56782 MW; 529E5B6692D0D7E9 CRC64;

Query Match 4.1%; Score 169.5; DB 1; Length 505;
 Best Local Similarity 22.6%; Pred. No. 0.00093;
 Matches 64; Conservative 59; Mismatches 127; Indels 33; Gaps 10;

QY 376 VELTEETFNATVWASDS---IYLFGAQMVAQVMAFLQSYIDVAVKLGKSTMLTRINC 431
 DB 28 LELTDNFSRISIDTSGAGLMLVEFFAPWCGHCKRLAPELEAATRIKQ--IYPLAKVDC 85
 QY 432 ADMSDVCTQNTVTEPFIKKYKKGPNVSYAGMLGTRDLKFTQLNRISYPVNTSIQEA 491
 DB 86 TANTNTCNKGVSGYPLTKIFRDGEAGAYDGPRTADGIYS--HLKKQAGPASPVLRTTEE 143
 QY 492 EETLSELYKDLILYSSVSLGLFSPMTAKADESEAGYLYLGTIGYSEEDVILLS 551
 DB 144 E-----FRKFLSDKDAIVGFFDSSFSEAHSEFLKAASNLRD---NYRFAHTNVESTIV 193
 QY 552 TKYASLPLALLAR--HTEGKIESIPLASTHAQ---DIYQITDALLLEMFPEITVENLP 605
 DB 194 NEEDDNGEGILLFRPSHLNKKFEEDKTVAYTEQMTSGKIKKFIQEMINFGICPHMTEDNKD 253

QY 606 SYFRLO-KPLLIFSD-----GTVPNQYKAILTVKOKYIDS 642
 DB 254 ---LIQKDLLILVYDVDEKNAKSGSNYMRNRMYWAKKFLDA 293

RESULT 7
 PDI ASPNG
 ID PDI ASPNG STANDARD; PRT; 515 AA.
 AC 012730;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Protein disulfide isomerase precursor (EC 5.3.4.1) (PDI).
 GN PDI OR PDI.
 OS Aspergillus niger.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5061;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBS 120.49 / N400;
 RX MEDLINE=97174110; PubMed=9021130;
 RA Ngiam C., Jeenes D.J., Archer D.B.;
 RT Isolation and characterisation of a gene encoding protein disulphide
 RL Curr. Genet. 31:133-138(1997).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL 3;
 RA Walplich S.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING
 CC DISULFIDE BONDS, MAY BE INVOLVED IN GLYCOSYLATION, PROLYL
 CC HYDROXYLATION AND TRIGGERCERIDE TRANSFER (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Rearrangement of both intrachain and
 CC interchain disulfide bonds in proteins to form the native
 CC structures.
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (By similarity).
 CC -1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X89797; CAA67332.1; -
 CC EMBL: X89449; CAA61619.1; -
 CC HSSP: P07237; IMEX.
 CC InterPro: IPR000886; ER_target.
 CC InterPro: IPR000063; ThioRed.
 CC Pfam: PF00085; ThioRed; 3.
 CC PRINTS: PRO0421; THIOREDOXIN.
 CC TIGRFAWS: TIGR01126; pdi_dom; 2.
 CC TIGRFAWS: TIGR01130; ER_PDI_fam; 1.
 CC PROSITE: PS00014; ER_TARGET; 1.
 CC PROSITE: PS00194; THIOREDOXIN; 2.
 KW Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.
 FT CHAIN 1 20 POTENTIAL.
 FT CHAIN 21 515 PROTEIN DISULFIDE ISOMERASE.
 FT DISULFID 54 57 REDOX-ACTIVE (BY SIMILARITY).
 FT DISULFID 369 392 REDOX-ACTIVE (BY SIMILARITY).
 FT SITE 512 515 PREVENT SECRETION FROM ER (POTENTIAL).
 SQ SEQUENCE 515 AA; 56291 MW; 2B005B8786400AD9 CRC64;

Query Match 4.1%; Score 169.5; DB 1; Length 515;
 Best Local Similarity 20.6%; Pred. No. 0.00095;
 Matches 111; Conservative 89; Mismatches 193; Indels 145; Gaps 27;

QY 15 VIMCIFYMPTVNSLPESLPQKIFSTLQGLLEINAEVRLQDYGISVAKVNCYKEISRY 74

```

Db 44 LVLAFFPWCCHKALAP-KY-----EEAATELKAKNIPLVVDCTAAE----- 87
QY 75 CGKEDLMKAYLFKNIIILREF-----PDTLFDVNAIVAHVIFALL--FSEKYYITN 125
Db 88 -----DLRSQVEEYPTLKIFRGVDSKSPQAGAROTESISYMKRQSLPAVSVNE-EN 141
QY 126 LEDLONINATLKGKANIIFSYRAIGIEPHRAVMEAGFYGGTYQFVLTETALLESIGS 185
Db 142 LEEITMKI-----VIGIYPSDDQETQAFEKYAESQRNYYLFAATDDAAIAIKSEV 195
QY 186 EDVEYAHLY--FPHCKLVLD--LTQOCRTIMEOPLTLNHLFKTKMAKPLLEVAEDP 241
Db 196 EOPSTIV-LYKDFDEKKAVYDGEIEQA-----HISWKSASTPLVGEIG--P 239
QY 242 QOVSTVHQLGLPLVFIYSQATYFADRTAEW-----VAMPLLG-KACV 285
Db 240 ETVSG-YIGAGVPLAYIFA-ETKEERERYTEDFKPIAQKHKAINIATIDAKMFAHAGN 297
QY 286 LLLLDLSLEVINIPD-----ANVYERABE--GVPEFLVHDVLLISHVENNMHIEEQ 339
Db 298 LNLDSQKPPARAIDPPANNAKYPIQAKELNADVEKEFI--QDVLGKVEPSISSEVP 354
QY 340 EDEDNDMEGPDIDVODDEVAETVFPRDKRKLPLELTVELTEETFNATYMASDSIYL--FY 397
Db 355 ESQ-----EGP-----VTV--VMAHSYKDIVDNDKDVLEFY 385
QY 398 AGQAVSAFPIQSYIDVAVKLGISTMTLITRINCADMSDVCYK-----ONV 443
Db 386 APWCHCKALAPKYDELALY-----ADHPDLAKVIAKIDATANDVPPPI 432
QY 444 TEFLPKYKKG--ENPVSYAGMLGTRKDLKPIOLNRSYPNITSIGAEFYLSGEL 499
Db 433 TGFPLRLRIPGAKDSPLEYSGSRTEVEDLANFVKENG-KHNDALNVAASEEQEGDV 489

```

```

RL Biochem. Biophys. Res. Commun. 178:679-685(1991).
RN [4]
RP SEQUENCE OF 26-43.
RC TISSUE=Brain, and pituitary;
RX MEDLINE=90208308; Pubmed=2181662;
RT Mobs C.V., Fink G., Pfaff D.W.;
RT "HIP-70": a protein induced by estrogen in the brain and LH-RH in the
RT pituitary";
RN Science 247:1477-1479(1990).
RP [5]
RP SEQUENCE OF 25-54; 258-269; 285-310; 347-350; 412-419 AND 434-463.
RC TISSUE=Liver;
RX MEDLINE=92041865; Pubmed=1657921;
RA Srivastava S.P., Chen N.O., Liu Y.X., Holtzman J.L.;
RT "purification and characterization of a new isozyme of thiol:protein-
RT disulfide oxidoreductase from rat hepatic microsomes. Relationship of
RT this isozyme to cytosolic phosphatidylinositol-specific phospholipase
RT C form 1A.";
RL J. Biol. Chem. 266:20337-20344(1991).
RN [6]
RP SEQUENCE OF 26-34; 174-193; 433-446 AND 448-458.
RX MEDLINE=92340568; Pubmed=1321829;
RA Urade R., Nasu M., Moriyama T., Wada K., Kito M.;
RT "Protein degradation by the phospholipase-specific phospholipase
RT C-alpha family from rat liver endoplasmic reticulum.";
RL J. Biol. Chem. 267:15152-15159(1992).
RN [7]
RP INHIBITION BY PHOSPHOLIPIDS.
RX MEDLINE=93050170; Pubmed=1330685;
RA Urade R., Kito M.;
RT "Inhibition by acidic phospholipids of protein degradation by ER-60
RT F6B5. Lett. 312:83-86(1992).
RL [8]
CC -1- CATALYTIC ACTIVITY: Rearrangement of both intrachain and
CC interchain disulfide bonds in proteins to form the native
CC structures.
CC -1- ENZYME REGULATION: SEEMS TO BE INHIBITED BY ACIDIC PHOSPHOLIPIDS.
CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
CC -1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
CC -1- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE A PHOSPHATIDYL-
CC INOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE TYPE I (PHOSPHOLIPASE
CC C-ALPHA) THEN WAS THOUGHT (REF.6 AND REF.7) TO BE A THIOL
CC PROTEASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X12355; CAA30916.1; -.
DR EMBL: D63378; BAA09695.1; -.
DR PIR: A28807; A28807.
DR PIR: A40095; A40095.
DR HSRP: P07237; IMEX.
DR MEROPS: C17.001; -.
DR InterPro: IPR000063; ThioRed.
DR Pfam: PF00085; ThioRed_2.
DR PRINTS: PR00421; THIOREDOXIN.
DR TIGRFAMS: TIGR01126; ptd_dom_2.
DR TIGRFAMS: TIGR01130; ER_PDI_fam; 1.
DR PROSITE: PS00194; THIOREDOXIN; 2.
DR Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.
FT SIGNAL 1 24
FT CHAIN 25 505
FT DISULFD 57 60
FT DISULFD 406 409
FT SITE 502 505
FT SITE 502 505
FT CONFLICT 1 13
FT FT

```


RT "Molecular structure of a yeast gene, PDI1, encoding protein
 RT disulfide isomerase that is essential for cell growth.";
 RL J. Biochem. 110:306-313(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91289690; PubMed-2063627;
 RA Scherens B., Dubois E., Messenguy F.;
 RT "Determination of the yeast YOL133 gene localized on
 RT chromosome III. Homology with the protein disulfide isomerase (PDI
 RT gene product) of other organisms.";
 RL Yeast 7:185-193(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91239586; PubMed-1840696;
 RA Imanita M., Miura T., Tachikawa H., Kaplan H.A., Lennarz W.J.,
 RA Mizunaga T.;
 RT "Glycosylation site binding protein and protein disulfide isomerase
 RT are identical and essential for cell viability in yeast.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:4453-4457(1991).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92105117; PubMed-1761554;
 RA Gienther R., Brauer C., Janetzky B., Foerster H.H., Ehbrecht I.M.,
 RA Lehle L., Kuentzel H.;
 RT "The Saccharomyces cerevisiae TRG1 gene is essential for growth and
 RT encodes a luminal endoplasmic reticulum glycoprotein involved in the
 RT maturation of vacuolar carboxypeptidase.";
 RL J. Biol. Chem. 266:24557-24563(1991).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92397595; PubMed-1523890;
 RA Scherens B., Messenguy F., Gigot D., Dubois E.;
 RT "The complete sequence of a 9,543 bp segment on the left arm of
 RT chromosome III reveals five open reading frames including glucokinase
 RT and the protein disulfide isomerase.";
 RL Yeast 8:577-586(1992).
 RN [6]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92104510; PubMed-1761235;
 RA Fargnhar R., Honey N., Murtant S.J., Bossier P., Schults L.,
 RA Montgomerie D., Ellis R.W., Freedman R.B., Tule M.F.;
 RT "Protein disulfide isomerase is essential for viability in
 RT Saccharomyces cerevisiae.";
 RL Gene 108:81-89(1991).
 CC -1- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING
 CC DISULFIDE BONDS. MAY BE INVOLVED IN GLYCOSYLATION, PROLYL
 CC HYDROXYLATION AND TRIGLYCERIDE TRANSFER.
 CC -1- CATALYTIC ACTIVITY: Rearrangement of both intrachain and
 CC interchain disulfide bonds in proteins to form the native
 CC structures.
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (potential).
 CC -1- PM: The N-terminus is blocked.
 CC -1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X52313; CAA36550.1; -
 DR EMBL: X57712; CAA40883.1; -
 DR EMBL: M62815; AAA34848.1; -
 DR EMBL: M76982; AAA35169.1; -
 DR EMBL: X59720; CAA42373.1; -
 DR EMBL: X54535; CAA38402.1; -
 DR EMBL: D00842; BAA00723.1; -
 DR PIR: JX0182; ISRYSS.
 DR PIR: S19372; S19372.
 DR PIR: S15050; S15050.
 DR PIR: S25349; S25349.

DR PIR: A41713; A41713.
 DR PIR: JS0634; JS0634.
 DR HSSP: P07237; IMEK.
 DR SGD: S0000548; PDI1.
 DR InterPro: IPR000886; ER_target.
 DR InterPro: IPR000063; ThioRed.
 DR Pfam: PF00085; ThioRed; 2.
 DR PRINTS: PR00421; THIAREDOXIN.
 DR TIGRFAMS: TIGR01126; pdi_dom; 2.
 DR TIGRFAMS: TIGR01130; ER_pdi_fam; 1.
 DR PROSITE: PS00194; ER_TARGET; 1.
 DR PROSITE: PS00194; THIAREDOXIN; 2.
 KW Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; signal.
 FT SIGNAL 1 28
 FT CHAIN 29 522
 FT DISULFID 61 64
 FT DISULFID 406 409
 FT CARBOHYD 82 82
 FT CARBOHYD 117 117
 FT CARBOHYD 155 155
 FT CARBOHYD 174 174
 FT CARBOHYD 425 425
 FT SITE 519 522
 FT CONFLICT 33 52
 FT CONFLICT 83 83
 FT CONFLICT 114 114
 FT CONFLICT 143 143
 FT CONFLICT 146 146
 FT CONFLICT 168 168
 FT CONFLICT 197 197
 FT CONFLICT 215 215
 FT CONFLICT 226 226
 FT CONFLICT 333 333
 FT CONFLICT 351 351
 FT CONFLICT 455 455
 FT CONFLICT 458 458
 FT CONFLICT 505 505
 SQ SEQUENCE 522 AA; 58227 MW; 690CF3E05D7F74C94 CnC64;
 Query Match 3.7%; Score 154.5; DB 1; Length 522;
 Best Local Similarity 20.8%; Pred. No. 0.0094;
 Matches 66; Conservative 67; Mismatches 126; Indels 58; Gaps 15;
 QY 358 VAETFERORRRKRLPLE-LTVELTEETFNATVWASQSYL-FAGQWQAVSMFLQSYDVA 415
 DB 16 LASSVFAQOEAVAPEDSAVAVKATDSFNEYIOSHDLVAEFPAWCGCHKNAPEYVKA 75
 QY 416 VLKGTSTMLTRINCADMSDVCTKQNTPEPIIKYKGE--NPVSAYAGMLGTDLKLF 473
 DB 76 ETLV-EKNITTLAQIDCTENQDLCMEHNTIPGFSLTIPNSVDNNSIDYEGPTAAIYOF 134
 QY 474 IQLNRISTPVNTISIQEAEVYISGLYKDLIYSSVVLGLFSPMKAKEDFSAGNYL 533
 DB 135 --MIQSQPA-VAVADDPAYLANETFTVPTVQSGKIDAEFNATF-----YSMAHHF 185
 QY 534 KGYVITGIYSEEDVLLSTKRYAASL--PALLARTECKIESIPLASTHAQDIYQITDA 591
 DB 186 NDYDVSAMADDDKSLIYLSANDEPVYVYNGKAD-----LIDA 226
 QY 592 LLEMPPE-ITVENLPISYRLOKPLILFSDGTNPQYKAILTL-----VKQYLDSETP 645
 DB 227 --DVEFKWLYQVEALPYFGEI-----DGSVFQYVESGLPLLYLNDEEELBEYKP 275
 QY 646 CWLNL--KNTPYGRGIL 660
 DB 276 LFTELAKKN---RGIM 288
 RESULT 11
 BS2_TRYBB STANDARD; PRT; 497 AA.
 ID BS2_TRYBB
 AC P12865;

DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE Bloodstream-specific protein 2 precursor.
 GN B22.
 OS Trypanosoma brucei brucei.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OX NCBI_TaxID=5702;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9001196; PubMed=2551375;
 RA Hsu M.P., Muchich M.L., Boothroyd J.C.;
 RT "A developmentally regulated gene of trypanosomes encodes a homologue
 of rat protein-disulfide isomerase and phosphoinositol-phospholipase
 C.";
 RL Biochemistry 28:6440-6446(1989).
 CC -1- FUNCTION: NOT KNOWN.
 CC -1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: J02865; AAA30168.1; -
 DR PIR: A32820; A32820.
 DR HSSP: P07237; 1MER.
 DR InterPro: IPR000063; ThioRed.
 DR Pfam: PF00085; ThioRed; 3.
 DR PRINTS: PR00421; THIOREDOXIN.
 DR TIGR: TIGR01126; pdl_dom. 2.
 DR PROSITE: PS00194; THIOREDOXIN; 2.
 KW Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.
 FT SIGNAL 1 14
 FT CHAIN 15 497
 FT DISULFID 48 51
 FT DISULFID 378 381
 FT CARBOHYD 30 30
 FT CARBOHYD 63 63
 FT CARBOHYD 85 85
 FT CARBOHYD 133 133
 FT CARBOHYD 154 154
 FT CARBOHYD 250 250
 FT CARBOHYD 278 278
 FT CARBOHYD 413 413
 FT CARBOHYD 465 465
 FT CARBOHYD 476 476
 FT CARBOHYD 482 482
 FT CARBOHYD 485 485
 FT CARBOHYD 488 488
 SQ SEQUENCE 497 AA; 55579 MW; 8A8EC4AD762B67EB CRC64;
 Query Match 3.5%; Score 147; DB 1; Length 497;
 Best Local Similarity 19.4%; Pred. No. 0.027;
 Matches 77; Conservative 76; Mismatches 158; Indels 86; Gaps 16;
 QY 375 TVLVEETFNATVMSD-STLVLYAG-----WQVSAVAFQSYDVAVKLGTS 422
 DB 21 SLTLRENNREITAKSEFLVRYVDTGCGQMLAPDEWMAA---NETIDNA----- 69
 QY 423 TMLLFRINCADSDVCTKWTEFFPIIKMKKENPVSVAIGTFLKFLDILNRISYP 482
 DB 70 --LLMGVDCDSQPELANANSIRGYPITLLFRNGKEAHNGAGTKDIIKIYKAN--VG 124
 QY 483 VNITSQDAEVEYISGELYDLIYSSVVLGFSPTMKAKDESPSAGNYLKYVITGAY 542
 DB 125 PAVTPASNAEEVYRAKEEDHV-----CVGLTANNSTSTTLABAQASFR----- 170
 QY 543 SEEDVLLSTKYVASLPALL-----LARHTGKIESIPLASTHADIVQITTDALLEN 595

DB 171 -----VSLKFEAEKLPPEDEKPEITYYVYRKGGKEYVDGMEYEKLTPELQISRVAF 223
 QY 596 FPEITVENLPSPYRLOKPL--LILFSDGTGVNPPQKRAITLVYKQYLDSEFTPCMLNKPT 653
 DB 224 GSEITPENQYYSVIRKPGWAMVKPNETASIELKES--LLEVGGKKRSHVYLVNVIKSH 282
 QY 654 PVGRGILIRAYFDLP--PLPLLVLYN-----LHSGGVFAFPASDQAIIIEENLVMLKKL 705
 DB 283 PVMRDF-----GVPEDAKYPALIAIHGANYLHSTAEEV-----VYRESLEKPILEF 328
 QY 706 EAG-LENHITLIPAQEMKPLPAYDFLSMIDATTSOR 741
 DB 329 AAGRVEPTIKSLPV-----PEVEYVDKTTIVATMOK 361
 RESULT 12
 PDIL_SCHPO
 ID PDIL_SCHPO STANDARD; PRT; 492 AA.
 AC 010057;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative protein disulfide isomerase ClF5.02 precursor (EC 5.3.4.1).
 GN SPAC1F5.02.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OX Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skellon J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Grynopre B.,
 RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Wellner S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leclaire V., Mottler S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garçon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC -1- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING
 DISULFIDE BONDS, MAY BE INVOLVED IN GLYCOSYLATION, PROXYL
 HYDROXYLATION AND TRIGLYCERIDE TRANSFER (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: Rearrangement of both intrachain and
 interchain disulfide bonds in proteins to form the native
 structure.
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (By similarity).
 CC -1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; 268136; CA92230.1; -.
 DR HSSP; P07237; IMK.
 DR InterPro; IPR000886; ER_target.
 DR InterPro; IPR000063; ThioRed.
 DR Pfam; PF00085; thioRed; 2.
 DR PRINTS; PR00421; THIOREDOXIN.
 DR TIGRfams; TIGR01126; pdl_dom; 2.
 DR TIGRfams; TIGR01130; ER_pdl_fam; 1.
 DR PROSITE; PS00014; ER_TARGET; 1.
 DR PROSITE; PS00194; THIOREDOXIN; 2.
 KW Hypothetical protein; Redox-active center; Isomerase;
 KM Endoplasmic reticulum; Repeat; signal.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 492 POTENTIAL.
 FT DISULFID 51 54 C1F5.02.
 FT DISULFID 385 388 REDOX-ACTIVE (BY SIMILARITY).
 FT CARBOHYD 161 161 REDOX-ACTIVE (BY SIMILARITY).
 FT CARBOHYD 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SITE 489 492 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 492 AA; 54880 MW; C50B59236566667 CRC64;
 Query Match 3.5%; Score 146.5; DB 1; Length 492;
 Best Local Similarity 19.3%; Pred. No. 0.029;
 Matches 103; Conservative 71; Mismatches 186; Indels 175; Gaps 20;
 QY 15 VINCIFVPTVNSLPDELSPQKFTSLQGLLENEAVRPLDQYGISVAKVNCVKE----- 69
 DB 41 VLMKFAVPCGCHKALP-----EYESADELEKDGISLVEVDCETEGSDICS 88
 QY 70 -----EISRCCKEK-DLMKAYLFKGNL-LAREPDTLFDVAAYAH 110
 DB 89 EYSIRGYPTLVNFKNGKQISQYSGPRKHDAVYKMKQLPTVKPDISKDTL----- 139
 QY 111 VLFALLFESEVKYITNLEPQNIENALKGKANIIFSYVRAIGIEPRRAVNEAGFYVGTYYQ 170
 DB 140 -----ENFVEKADLAVVAFKDKLNDTY-----EVAEYKDKDFV----- 176
 QY 171 FVLTEIALLESIGSEDEVYALHYFHCKLVDLTQGRRTLMEOPLTTLNTHLEIKTWK 230
 DB 177 FAASDCKELAKSLGS--NFGIVAF-----TKDAQDSQKLVYTGDMWPASIADELTGVSS 229
 QY 221 APLTEVADPQVSTVHLQGLPLVFT----- 258
 DB 230 IPLADEL---NQMTFGKYOQSGPLGIIFYNSTESRDELYDFOPLAKKYQDTLRFAPLD 286
 QY 259 -----VSQQAITYEADRTAEVAVRILGKAGVLLLRDSLEVNIPQDANVYFKRAEGV 312
 DB 287 AVRIGAVAKQANVESD-----WPAF-----VIANLSMKLYPPRP--TTELTAKAMTFEV 333
 QY 313 PVEFLVLDVLLISHVENNMHIEIOEDENDEMGPPIDVQDDVAETVFRDKRKPL 372
 DB 334 G-----DFVQDKLQPKIKSQPIPEQED----- 356
 QY 373 ELTYVELTEETENATVM--ASDSIVLFYAGQAVSMAFLQSYIDAVVKLGKSTMLLTRIN 430
 DB 357 --LVVLVADNPDIDYMDETKVDLYEFVAPWCGCHKKNLAPTYEKLAEEVSDSNVAVAKID 414
 QY 431 CADMSDQVTKNQTEPFIKKMYKGE--NPVSYAGMLGKDKLQKLTQNRISYPV 483
 DB 415 ATE-NDISV--SISGFTIMEFKANDKAVPARYEGDRLELDSAFIDKHASEPI 466
 RESULT 13
 PDAA_RAT STANDARD: PRT: 643 AA.
 AC P36659;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Protein disulfide isomerase A4 precursor (EC 5.3.4.1) (Protein Erp-72)
 DE (Erp72) (Calcium-binding protein 2) (Cabp2).
 DE -----

GN PDIA4 OR CABP2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar; TISSUE=Liver;
 RX MEDLINE=93238767; PubMed=8477750;
 RA Van P.N., Rupp K., Lampen A., Soelling H.-D.;
 RT "Cabp2 is a rat homolog of Erp72 with proteindisulfide isomerase
 RT activity";
 RL Eur. J Biochem. 213:789-795(1993).
 CC -1 CATALYTIC ACTIVITY: Rearrangement of both intrachain and
 CC interchain disulfide bonds in proteins to form the native
 CC structures.
 CC -1 SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
 CC -1 INDUCTION: UPON GLUCOSE STARVATION, AS WELL AS TREATMENT WITH
 CC TUNICAMYCIN.
 CC -1 PTM: O-GLYCOSYLATED.
 CC -1 SIMILARITY: CONTAINS 3 THIOREDOXIN DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M86870; AAA19217.1; -.
 DR HSSP; P07237; IMK.
 DR MEROPS; C17.002; -.
 DR InterPro; IPR000886; ER_target.
 DR InterPro; IPR000063; ThioRed.
 DR Pfam; PF00085; thioRed; 3.
 DR PRINTS; PR00421; THIOREDOXIN.
 DR TIGRfams; TIGR01126; pdl_dom; 3.
 DR TIGRfams; TIGR01130; ER_pdl_fam; 1.
 DR PROSITE; PS00014; ER_TARGET; 1.
 DR PROSITE; PS00194; THIOREDOXIN; 3.
 KW Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; signal;
 KW Glycoprotein; Calcium-binding.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 643 PROTEIN DISULFIDE ISOMERASE A4.
 FT DISULFID 89 92 REDOX-ACTIVE (BY SIMILARITY).
 FT DISULFID 204 207 REDOX-ACTIVE (BY SIMILARITY).
 FT DISULFID 553 556 REDOX-ACTIVE (BY SIMILARITY).
 FT SITE 640 643 PREVENT SECRETION FROM ER.
 SQ SEQUENCE 643 AA; 72806 MW; 856118663A4DE41 CRC64;

Query Match 3.4%; Score 142; DB 1; Length 643;
 Best Local Similarity 22.8%; Pred. No. 0.086;
 Matches 115; Conservative 76; Mismatches 201; Indels 112; Gaps 27;

QY 15 VINCIFVPTVNSLPDELSPQKFTSLQGLLENEAVRPLDQYGISVAKVNCVKEISKY 74
 DB 194 ILVVEYVAPWCGCHKLAPE-----YEKAKELSKRSPP-----IPLAKVDAT----- 236
 QY 75 CKEKEDKMAVLFKGNILLREPTDLPVNV-----AIVAHVLFALLFESEVKYITNED 128
 DB 237 --EQDTLAKRPVSGPTLKIFRKGRPPYNGPREKYGIADVWVDESGSPSEKILL----- 290
 QY 129 LQNIENALK-KRANIIIFYVRAIGIEPRRAVNEAGFYVGTYYQF--VLTEIA--LLES1 183
 DB 291 LKQVGEFLKDGDDVILVFGQGVDPGYLQYQDANATLRBEDYKFNHTSTELAKFLKXSL 350
 QY 184 GSEDEVYALHYFHCKLVL-----DLTQGRRTLMEOPLTT--LNHLFITMKAPLL- 234
 DB 351 G-----KLIVMQPEKFKQSKYEPRMHNVADVQSTEASIKDYVVKHDLPLVG 396

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:11:22 ; Search time 36 Seconds
(without alignments)
4618.889 Million cell updates/sec

Title: US-09-847-046-2
Perfect score: 4142
Sequence: 1 MFGSFNVFRVIGISFVIMCIF.....FKEAKSFRDRKELGCSKVN 807

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
1: SPREMBL_21:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.rodent:*
13: sp.virus:*
14: sp.vertibrate:*
15: sp.unclassified:*
16: sp.rvirts:*
17: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------|--------------------|
| 1 | 4025 | 97.2 | 806 | 4 Q9P2K2 | Q9P2K2 homo sapien |
| 2 | 1839 | 44.4 | 357 | 4 Q9H9W7 | Q9H9W7 homo sapien |
| 3 | 1423 | 34.4 | 354 | 11 Q9CS82 | Q9CS82 mus musculu |
| 4 | 815 | 19.7 | 225 | 11 Q8R2W8 | Q8R2W8 mus musculu |
| 5 | 184.5 | 4.5 | 502 | 3 Q74568 | Q74568 trichoderma |
| 6 | 182 | 4.4 | 546 | 10 Q9C818 | Q9C818 arabidopsis |
| 7 | 178 | 4.3 | 519 | 10 Q9LW75 | Q9LW75 arabidopsis |
| 8 | 176 | 4.2 | 489 | 5 Q76945 | Q76945 fasciola he |
| 9 | 169.5 | 4.1 | 482 | 5 Q26593 | Q26593 schistosoma |
| 10 | 168.5 | 4.0 | 505 | 11 Q9JLF6 | Q9JLF6 mus musculu |
| 11 | 166.5 | 4.0 | 505 | 11 Q91Z81 | Q91Z81 cricetus |
| 12 | 162 | 3.9 | 577 | 16 Q8R6N2 | Q8R6N2 thermoaer |
| 13 | 158 | 3.8 | 369 | 13 Q91815 | Q91815 fugu rubrip |
| 14 | 154 | 3.7 | 494 | 5 Q9GPH2 | Q9GPH2 bombyx mori |
| 15 | 150.5 | 3.6 | 517 | 3 Q9C128 | Q9C128 pichia past |
| 16 | 150 | 3.6 | 579 | 10 Q8VX13 | Q8VX13 arabidopsis |

| | | | | | |
|----|-------|-----|------|-----------|---------------------|
| 17 | 148 | 3.6 | 566 | 10 Q9SV44 | Q9SV44 arabidopsis |
| 18 | 147.5 | 3.6 | 483 | 5 Q9GR12 | Q9GR12 plasmodium |
| 19 | 144.5 | 3.5 | 359 | 10 P93358 | P93358 nicotiana t |
| 20 | 143.5 | 3.5 | 2471 | 12 Q9YTK3 | Q9YTK3 ateline her |
| 21 | 141 | 3.4 | 436 | 5 Q01492 | Q01492 caenorhabdi |
| 22 | 138.5 | 3.3 | 416 | 5 Q9VYV3 | Q9VYV3 drosophila |
| 23 | 138.5 | 3.3 | 487 | 3 Q96VF3 | Q96VF3 ussilaago ma |
| 24 | 137 | 3.3 | 366 | 10 Q93V71 | Q93V71 oryza sativ |
| 25 | 136 | 3.3 | 930 | 16 Q9CUX4 | Q9CUX4 pasteurella |
| 26 | 135 | 3.3 | 497 | 5 Q76191 | Q76191 dirofilaria |
| 27 | 134.5 | 3.2 | 440 | 10 Q48773 | Q48773 arabidopsis |
| 28 | 133 | 3.2 | 572 | 5 Q44508 | Q44508 caenorhabdi |
| 29 | 132.5 | 3.2 | 2795 | 2 Q9RNB1 | Q9RNB1 microcystis |
| 30 | 131.5 | 3.2 | 574 | 5 P91442 | P91442 caenorhabdi |
| 31 | 130.5 | 3.2 | 363 | 3 Q13704 | Q13704 schizosacch |
| 32 | 130.5 | 3.2 | 485 | 5 Q17967 | Q17967 caenorhabdi |
| 33 | 130.5 | 3.2 | 509 | 11 Q8R402 | Q8R402 cricetus |
| 34 | 128.5 | 3.1 | 2149 | 10 Q9M3D3 | Q9M3D3 arabidopsis |
| 35 | 128 | 3.1 | 977 | 10 Q9X114 | Q9X114 arabidopsis |
| 36 | 127.5 | 3.1 | 480 | 5 Q96460 | Q96460 schistosoma |
| 37 | 127.5 | 3.1 | 488 | 5 Q17908 | Q17908 caenorhabdi |
| 38 | 127.5 | 3.1 | 489 | 5 Q9TWZ1 | Q9TWZ1 drosophila |
| 39 | 127.5 | 3.1 | 509 | 11 Q922C8 | Q922C8 mus musculu |
| 40 | 127 | 3.1 | 443 | 10 Q9MAU6 | Q9MAU6 arabidopsis |
| 41 | 126.5 | 3.1 | 1014 | 2 Q8RN21 | Q8RN21 campylobact |
| 42 | 126.5 | 3.1 | 1014 | 2 Q8R1Y7 | Q8R1Y7 campylobact |
| 43 | 126.5 | 3.1 | 1214 | 17 Q8U3H2 | Q8U3H2 pyrococcus |
| 44 | 126 | 3.0 | 1042 | 1 Q9C4Y4 | Q9C4Y4 sulfobolus |
| 45 | 125.5 | 3.0 | 371 | 10 Q94212 | Q94212 oryza sativ |

ALIGNMENTS

| | | | | | |
|---|---|------|-----|-----|--|
| RESULT 1 | | | | | |
| Q9P2K2 | PRELIMINARY; | PRT; | 806 | AA. | |
| AC | Q9P2K2: | | | | |
| DT | 01-OCT-2000 (TREMBLrel. 15, Created) | | | | |
| DT | 01-OCT-2000 (TREMBLrel. 15, Last sequence update) | | | | |
| DT | 01-DEC-2001 (TREMBLrel. 19, Last annotation update) | | | | |
| DE | KIAA1344 protein (Fragment). | | | | |
| GN | KIAA1344. | | | | |
| OS | Homo sapiens (Human). | | | | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | | | |
| OX | NCBI_TaxID=9606; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | TISSUE=BRIN; | | | | |
| RX | MEDLINE=20181126; PubMed=10718198; | | | | |
| RA | Nagase T., Kikuno R., Ishikawa K., Hirosewa M., Ohara O.; | | | | |
| RT | "Prediction of the coding sequences of unidentified human genes. XVI. | | | | |
| RT | The complete sequences of 150 new cDNA clones from brain which code | | | | |
| RT | for large proteins in vitro." | | | | |
| RL | DNA Res. 7:65-73(2000). | | | | |
| DR | EMBL: AB037765; BAA92582.1; - | | | | |
| DR | Interpro: IPR000063; Thioired. | | | | |
| FT | NON_TER | | | | |
| FT | SEQUENCE 806 AA: 91407 MW; B9A4847F928662E3 CRC64; | | | | |
| Query Match | | | | | |
| Best Local Similarity 97.2%; Score 4025; DB 4; Length 806; | | | | | |
| Matches 788; Conservative 0; Mismatches 0; Indels 18; Gaps 1; | | | | | |
| QY | 20 FYMPVNSLPFLSPQKVFSTLPG-----LEELNEAVRPLDYGVISV 61 | | | | |
| DB | 1 FYMPVNSLPFLSPQKVFSTLPGKASLAFQADSPRTSVLEELNEAVRPLDYGVISV 60 | | | | |
| QY | 62 AKVNCVKKEISRYCKEKDLMKAYLFKGINLLRPPDTLFFVNAIVAVLFFSEVK 121 | | | | |
| DB | 61 AKVNCVKKEISRYCKEKDLMKAYLFKGINLLRPPDTLFFVNAIVAVLFFSEVK 120 | | | | |

QY 122 YTNLEDLQNIENALKNANIIESYVRAIGIPEHRAVMEAGFYVGTYYQVLTTEIALLE 181
 DB 121 YTNLEDLQNIENALKNANIIESYVRAIGIPEHRAVMEAGFYVGTYYQVLTTEIALLE 180
 QY 182 SISEDEYVHAFYFHKLVLDLTOOCRTLMEOPLTTLNHLFIKMKAPLITLVEADP 241
 DB 181 SISEDEYVHAFYFHKLVLDLTOOCRTLMEOPLTTLNHLFIKMKAPLITLVEADP 240
 QY 242 QQSVTHVLOGLPLVIVSOQATEADRTAEVWAMRLKAGVLLLRSLSEVNIQDA 301
 DB 241 QQSVTHVLOGLPLVIVSOQATEADRTAEVWAMRLKAGVLLLRSLSEVNIQDA 300
 QY 302 NVVFKRAEEVPEVFLVHDVLLIISVENNMHIEEIQDEBNDMEGPDIDVDEVAET 361
 DB 301 NVVFKRAEEVPEVFLVHDVLLIISVENNMHIEEIQDEBNDMEGPDIDVDEVAET 360
 QY 362 VFDPRKRLPLELTVLEETFPNATVMAISIVFYAGMVAISAFLOSIVDAVVKLGT 421
 DB 361 VFDPRKRLPLELTVLEETFPNATVMAISIVFYAGMVAISAFLOSIVDAVVKLGT 420
 QY 422 STMILTRINCADMSDVCTKONVTEFPILIKMYKGENPVSYAGMLGTDLKFTOLNRISY 481
 DB 421 STMILTRINCADMSDVCTKONVTEFPILIKMYKGENPVSYAGMLGTDLKFTOLNRISY 480
 QY 482 PVNITSIQEABEYISGLYKDLIIYSVSVGLFSPITMKAKEDFSAGNYLKYVTIGI 541
 DB 481 PVNITSIQEABEYISGLYKDLIIYSVSVGLFSPITMKAKEDFSAGNYLKYVTIGI 540
 QY 542 YSEEDVLLSTKYAASIPALLARHTEGKIESIPLASTHODIYOITDALLMEFPETIV 601
 DB 541 YSEEDVLLSTKYAASIPALLARHTEGKIESIPLASTHODIYOITDALLMEFPETIV 600
 QY 602 ENLPSTFRLQKPLILFSDGTVPNOYKKAITLVKORYLDSFPWCMLNKNTPVGRGILR 661
 DB 601 ENLPSTFRLQKPLILFSDGTVPNOYKKAITLVKORYLDSFPWCMLNKNTPVGRGILR 660
 QY 662 AYDPLPLPLVLYVNHSGGVAFAPSDAIIIEENLVMLKLEAGLENNHTITLPAQEW 721
 DB 661 AYDPLPLPLVLYVNHSGGVAFAPSDAIIIEENLVMLKLEAGLENNHTITLPAQEW 720
 QY 722 KPPLPAYDFLSMIDATSOGRTRKVPKCKMETVQENDKQCHEDKSAVRKEPIETLRKH 781
 DB 721 KPPLPAYDFLSMIDATSOGRTRKVPKCKMETVQENDKQCHEDKSAVRKEPIETLRKH 780
 QY 782 WNRSMNFKEAKESFRDKELGCSKVN 807
 DB 781 WNRSMNFKEAKESFRDKELGCSKVN 806

RESULT 2
 ID 09H9W7 PRELIMINARY; PRT; 357 AA.
 AC 09H9W7;
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE CDNA FLJ12501 f1s, clone NTMRM200161.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN NCB1;TaxID=9606;
 RP SEQUENCE FROM N.A.
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
 RA Matsunawa H., Ishii S., Kawai Y., Saio K., Yamamoto J., Wakamatsu A.,
 RA Nakamura Y., Naganari K., Masuno Y., Sasaki N.;
 RT "NEO human cDNA sequencing project."
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK022563; BAB14101.1; -
 SQ SEQUENCE 357 AA; 40469 MW; A7FCD05C2D2E560 CRC64;

Query Match 44.4%; Score 1839; DB 4; Length 357;

Best Local Similarity 99.7%; Pred. No. 2,3e-123;
 Matches 356; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 451 MYKKGENVSYAGMLGTGKDLKFTQLNRISYPVNITSIQEABEYISGLYKDLIIYSVS 510
 DB 1 MYKKGENVSYAGMLGTGKDLKFTQLNRISYPVNITSIQEABEYISGLYKDLIIYSVS 60
 QY 511 VLGIFSPITMKAKEDFSAGNYLKYVITGISEEDVLLSTKYAASIPALLARHTEG 570
 DB 61 VLGIFSPITMKAKEDFSAGNYLKYVITGISEEDVLLSTKYAASIPALLARHTEG 120
 QY 571 IESTPLASTHODIYOITDALLMEFPETIVENLPSTFRLQKPLILFSDGTVPNOYKKA 630
 DB 121 IESTPLASTHODIYOITDALLMEFPETIVENLPSTFRLQKPLILFSDGTVPNOYKKA 180
 QY 631 ILTVKORYLDSFPWCMLNKNTPVGRGILRAYFDPPLPLVLYVNHSGGVAFAPSD 690
 DB 181 ILTVKORYLDSFPWCMLNKNTPVGRGILRAYFDPPLPLVLYVNHSGGVAFAPSD 240
 QY 691 QAIIEENLVMLKLEAGLENNHTITLPAQEWKPPPLPAYDFLSMIDATSOGRTRKVPKM 750
 DB 241 QAIIEENLVMLKLEAGLENNHTITLPAQEWKPPPLPAYDFLSMIDATSOGRTRKVPKM 300
 QY 751 KETDVOENDKQCHEDKSAVRKEPIETLRKHNRSMNFKEAKESFRDKELGCSKVN 807
 DB 301 KETDVOENDKQCHEDKSAVRKEPIETLRKHNRSMNFKEAKESFRDKELGCSKVN 357

RESULT 3
 ID 09C882 PRELIMINARY; PRT; 354 AA.
 AC 09C882;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE 5730420B22RIK protein (Fragment).
 GN 5730420B22RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN NCB1;TaxID=10090;
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Glassi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsch G.,
 RA Blake J., Boffelli D., Bojunga N., Carrincci P., de Bonaldo M.F.,
 RA Brownstein M.J., Buit C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Hayashizaki Y., Yoshida K., Hasegawa Y., Kawai J., Kohsaki S.,
 RA Wyshaw-Boris A.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL; AK017582; BAB30819.1; -
 DR MGD; MGI:1917811; 5730420B22RIK.
 FT NON_TER 354
 SQ SEQUENCE 354 AA; 39924 MW; 42B75F03910AEC75 CRC64;

Query Match 34.4%; Score 1423; DB 11; Length 354;
 Best Local Similarity 77.6%; Pred. No. 1.3e-97;
 Matches 274; Conservative 34; Mismatches 27; Indels 18; Gaps 1;


```

Db      484 SESKASE 490

RESULT 6
09C818 ID 09C818 PRELIMINARY; PRT; 546 AA.
AC 09C818;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Protein disulfide isomerase precursor, putative, 72379-69727.
GN F19K6.17.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxId=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Bueller E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremetska I., Kurtz D.B., Kwan A., Lam B.,
RA Lang-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Malt R., Marzilli A.,
RA Miltner J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana."
RL Nature 408:816-820(2000).
DR EMBL: AC037424; ANG51554.1; -
DR HSSP: P07237; IMEX.
DR InterPro: IPR000886; ER_target.
DR Pfam: PF00085; ThioRed.
DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
KW Isomerase. 546 AA; 6116 MW; E7B206FEF239270C CRC64;
SQ
SEQUENCE 546 AA; 6116 MW; E7B206FEF239270C CRC64;

Query Match 4.4%; Score 182; DB 10; Length 546;
Best Local Similarity 20.9%; Pred. No. 0.00015;
Matches 111; Conservative 93; Mismatches 180; Indels 148; Gaps 25;

QY 317 LVLDVLDLIISHVENNMHIEIOEDENDME---GPDIVODD----EVAETVFRDKR 368
DB 19 LILSLTIIIAVSSPDSSNESNEPGFSDLDQLAVDEQIQEERPEQSEAEFV--SKAQ 76
QY 369 KLPLELVELTEETFNATWASDSIYLFYAGQAVSMAPLOSTIDVAVKLKGFTSTMLLT 427
DB 77 RYLELVNGDTKRVLDGNEFV---VVLGYAPWCARSALMPRAEATATALKEIGSSVLA 133
QY 428 RINCADMSDVCTKQNTVEPIIKMKKGENPVSYAGMLGTDLKLTQLNRISYPVNTS 487
DB 134 KIDGDSYKIASLELEKGFPTLLFPVNGTS-LTYNGSSMEDIVINWQKTGAPITLNT 192
QY 488 IOEAEEYLSGELYKDLILYSSVGLF-----SPTWKTAKED--FSE----- 528
DB 193 VDEAPRFL--DKYHTF-----VLGLFEKFESEHNEFVKAASDEIQFIETRSDVA 243
QY 529 -----AGNYLKGFT-----VITGYSEEDVL--LSTKY----- 554
DB 244 KLLPDLKSNMVFILGVDEAEERYTVDSYKMKELIEFLGSKNKFLLTKLTETNTWVY 303

```

```

QY 555 --AASIPALLARHTEKIESIPLASTHADIQIITDALLEMFPEITVENLDSYFLQK 612
DB 304 SSPVKQLQVLMFSKADDFQKLAQPL-----EDIAKFKSKLMFIYDITNEN-----LAM 352
QY 613 PLILTFSDGTVPNQYKKAILLVKKYLDFTFCWMLNKN--IPVSGILRAYF--DPLP 668
DB 353 PFLIFGIEAGNTVVAAPFNNNLSKYLSDSPNSIEEFCGSLAHGTVSRYSRSEPPV 412
QY 669 PLPLVLVNLHSGGVAFPPSDAII--EENLVL-----WLKRLKA-----G 708
DB 413 DNEASAVIVY--VGKTF-----DGLVLSRENVILLEVTTPCVCNCEALSKQIEKLAKHFKG 466
QY 709 LENHTITLPAQEKMPPLPAVDFLSMDATSQKTRKVRPKMKTVDYQENDK 760
DB 467 FENLVY-----FARIDASANE-----HTKQLQVDDK 490

RESULT 7
Q9LW75 ID Q9LW75 PRELIMINARY; PRT; 519 AA.
AC Q9LW75;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE dbj|BAAB1871.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxId=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones."
RL DNA Res. 7:131-135(2000).
DR EMBL: AB012247; BAB02677.1; -
DR HSSP: P07237; IMEX.
DR InterPro: IPR000886; ER_target.
DR Pfam: PF00085; ThioRed.
DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
SQ SEQUENCE 519 AA; 57859 MW; 40A4AB1C7CD2F123 CRC64;

Query Match 4.3%; Score 178; DB 10; Length 519;
Best Local Similarity 21.0%; Pred. No. 0.00027;
Matches 114; Conservative 102; Mismatches 196; Indels 132; Gaps 27;

QY 315 EFLVLDVLDLIISHV-----ENNMHIEIOEDENDME---GPDIVODD----EVAET 361
DB 10 KESILTFPLLSFLPLFVARSSDVAVBAGSEELDDLEQLAVDEQIQEERPEQSEAE 69
QY 362 VFDRKKRKLPLELVELTEETFNATWASDSI-VLFYAGQAVSMAPLOSTIDVAVKLG 420
DB 70 VSKAQF-----LVLELNGDNTKRLDNGEYVAVLGYAPWCARSALMPRAEATATDKE 123
QY 421 T-STMLLTRINCADMSDVCTKQNTVEPIIKMKKGENPVSYAGMLGTDLKLTQLNRI 479
DB 124 IGSYVLMARKIDGRYSKVASQLEIKGFPTLLFVNGTSQ-SYNGGFSSEIYIYWQKTG 182
QY 480 SYPVNTISQEAEEYL-----SG--ELYK-----DLIYSSVVL 512
DB 183 ASTIKLDIVDEASGLFKHHHTLILGLFEKSESDSGHDEYVKAASLDNEIQFVETSSIDVA 242

```


QY 513 GLFSPMTAKEDFEAGNYLKGVIYIT-----GIYSEEDVL--LSTFYAASLPAL 561
 Db 243 KLFEPNLKT-----NNVEYGLVKTETAKTYSIDGPOQAKIYEFPLSNKFP----- 288
 QY 562 LIARHTEG-----KIESIPLASTH-----ADIVQIITDALLEPEITVEN 603
 Db 289 LVTKLTESVTVAVSSPVKIQVAFVSKTDDESLAQPLEDIARKFKSKIMLIYIDISNEN 348
 QY 604 LPSYRLQKPLLILFSDGTVPQYKKAILLTVKQYLDSPFCWMLNKTPTVG--RGLLR 661
 Db 349 -----LAMPFLTLFGIEDAKKTVAAPFNNLNLSKYLSDSPSPSIEFCGLAHGTVS 402
 QY 662 AVF--DPLPP--LPLVLVNLHSGGVAFPSDOAI--ENNVLMKLKLE--AGLEN 711
 Db 403 AVYKQSPIDPQNMASVAVV-----GRTP-----DEVVLASSENVLLLEVKLQSHKFGFEN 453
 QY 712 HI--TILPAQEKPPPLPAYDPLSMIDATVSQGRTRKVRKCKMETVDVQENDKQHEDKSA- 768
 Db 454 LVFARIDASANEHPKLTVDYPTIILYKTGE-----KENPLKLTSSAKDMAVL 503
 QY 769 VRKE 772
 Db 504 INKE 507

RESULT 8

ID 076945 PRELIMINARY: PRT: 489 AA.
 AC 076945:
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Protein disulphide isomerase.
 GN PDI GENE.
 OS Fasciola hepatica (liver fluke).
 CC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
 OC Echinostomidae; Echinostomata; Fasciolidae; Fasciola.
 OX NCBI_TaxID=6192;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Salazar M., Casals R., Diaz A., Martin J.M., Boga J.A., Parra F.;
 RT "Cloning and heterologous expression of Fasciola hepatica putative
 RT protein disulphide isomerase.";
 RL Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ225804; CAAL2644.1;
 DR HSSP: P07237; IMEX.
 DR InterPro: IPR000886; ER_target.
 DR InterPro: IPR000063; ThioRed.
 DR Pfam: PF00085; ThioRed; 3.
 DR PRINTS: PR00421; THIOREDOXIN.
 DR TIGRFAMS: TIGR01130; ER_PDI_fam; 1.
 DR TIGRFAMS: TIGR01126; pdi_dom; 2.
 DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
 DR PROSITE: PS00194; THIOREDOXIN; 2.
 DR Isomerase; Redox-active center.
 SQ SEQUENCE 489 AA; 55222 MW; 047AF15027C872A4 CRC64;

Query Match 4.2%; Score 176; DB 5; Length 489;
 Best Local Similarity 23.2%; Pred. No. 0.00035;
 Matches 65; Conservative 60; Mismatches 99; Indels 56; Gaps 12;

QY 376 VELTEETFNATVMSD-STVLFYACQMAVSMALQSYIDVAVKLKGT--STMTLTRINCAD 433
 Db 31 VELTEETDEDEKKKEFAVMYAPMCHGCKMAKPEYARAAQOLKEEGSDIMIAKVDAIQ 90
 QY 434 WSDVCTKQNVTEPIIKMYKKGENVSYAGMLGTRDLKFLQJNLRISYVNTS--IOEAE 492
 Db 91 HSKLAKSHNVTVGPTLKFYKSGV-WLDYTGQGTQKEIYMWIK-RKVSAPVSLSTLSVQ 148
 QY 493 EYLSBELKDLILYSSVSLGLFSPMTAKEDFEAGNYLKGVIYITGIYSEEDVLLST 552
 Db 149 QLVDEK--DIVV-----IAFAEESNEELKOLLE--AVASVYDKREGEFVSS 190

QY 553 KYA-----ASPLALLARHTEGKIESIPLASTHAQDIVQIITDALLEPEITVENLP 605
 Db 191 KQAFDHYKIDSKSRVLEFKKFDGR-----ADFGELREALIEFMOKETI----- 236
 QY 606 SYFRLQKPLLILFSDGTVPQYKKAILLTVKQYLDSPFCWMLNKTPTVG--RGLLR 661
 Db 237 -----PLVETQETASAVFCSAI-----RKHVVSFVP 264

RESULT 9

ID 026593 PRELIMINARY: PRT: 482 AA.
 AC 026593:
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Protein disulphide isomerase homologue precursor.
 OS Schistosoma mansoni (Blood fluke).
 CC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidae;
 OC Schistosomatidae; Schistosomatidae; Schistosoma.
 OX NCBI_TaxID=6183;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LIBERIA;
 RX MEDLINE=94359529; PubMed=8078516;
 RT Finken M., Sobek A., Symmons P., Kunz W.;
 RT "Characterization of the complete protein disulphide isomerase gene of
 RT Schistosoma mansoni and identification of the tissues of its
 RT expression.";
 RT Mol. Biochem. Parasitol. 64:135-144 (1994).
 DR EMBL: Z22933; CA80520.1;
 DR HSSP: P07237; IMEX.
 DR InterPro: IPR000886; ER_target.
 DR InterPro: IPR000063; ThioRed.
 DR Pfam: PF00085; ThioRed; 2.
 DR PRINTS: PR00421; THIOREDOXIN.
 DR TIGRFAMS: TIGR01130; ER_PDI_fam; 1.
 DR TIGRFAMS: TIGR01126; pdi_dom; 2.
 DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
 DR PROSITE: PS00194; THIOREDOXIN; 2.
 DR Isomerase; Redox-active center; Signal.
 FT SIGNAL 1
 FT SIGNAL 28
 SQ SEQUENCE 482 AA; 54154 MW; EA2DE2E2BB209658 CRC64;

Query Match 4.1%; Score 169.5; DB 5; Length 482;
 Best Local Similarity 22.3%; Pred. No. 0.00099;
 Matches 117; Conservative 70; Mismatches 170; Indels 167; Gaps 29;

QY 16 INCIEFMPYVNSLPELSPQRYESTLOPGLLELNEAVRPLQDYG--TSVAKVNC-VKEEIS 72
 Db 43 VLVEFYAPMCHGCKKALAP-----EYSEAKKIKLEKSGSLIKLAKVDATVEEELA 90
 QY 73 RYCGKCKDKMLKAVLEKGNILLRFPPTDL--FDVNAVAVHVLFPALFSE--VKYITNLED 128
 Db 91 LKHG-EKGYPTLKFFR-----NEQPIDFLGERSDAIVN--WCLKRSKPSVEYIDLS 141
 QY 129 LQNIENALSKANI--IFSYYR--ATGIRPHRYVM-----EAGRYVG-----TTYQFVLE 174
 Db 142 CKQFID--KANIALTLGFTIKDTSIDLADFEKVADELDDAGAIANSSEILTEYITQT 197
 QY 175 TEIALLSIGSDVEYAHLYFFHCKLVLDLTQCCRTLMEOPLTTLNHLFIKTMAPPL 234
 Db 198 PKIVLEFKNDENRVEY-----GGTLENLKHFIQVESVPLV 233
 QY 235 TEVAEDPQOVSTVHLQGLP-----LVFIVSQATYV--ADRTAEVNAVRLGKAGVLL 288
 Db 234 SEESQ-----KTAGVYFGSPIDKHLIVFLSKSTSDHSDLVDKLLE--VAHQFGKHLVIV 286
 QY 289 LKDSLEVINIPQDANVYFKRAEGSVPEVFLVLDVLLIIVENKHTIE----- 337
 Db 287 -----DVD-----VENNLKLVLEFFGLSKNDAPT 309
 QY 338 ---IQDEEDNDMEGPD-----IDVDDEVAETVFRKRKKPLLELIVE-----L 378

```

DB 310 YRIELGEETTKYKPDNDYSVSAMSEDFVORTI--DGKVK-PFLMSEIPSDDTGAVVL 366
OY 379 TEETFNATV--MASDSIVLFYAGQAVSMAPLOSYIVAVKLTGSTMLLRICADMSD 436
DB 367 VGKRYNVDKSDKDVAVKLAAPWCCHKALAPVWDELGEFFKSDVIAK-----MD 419
OY 437 VCTKQ---NTEEPPIKMY-KKGENPVSYAGMLGTDLKLFIO 475
DB 420 ATVNEVEDLKVTSPFLTKFYPKNSEVIDYTGDRSFALKKFE 463

RESULT 10
O99LF6 PRELIMINARY: PRT; 505 AA.
ID O99LF6;
AC O99LF6;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Similar to ER-60 protease.
GN GRP58.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003285; AAH03285.1; -.
DR HSSP; P07237; IMEX.
DR MGD; MGI:95834; GIP58.
DR InterPro; IPR000063; ThioRed.
DR Pfam; PF00085; thioRed; 2.
DR PRINTS; PR00421; THIOREDOXIN.
DR TIGRFSMS; TIGR01130; ER_PDI.fam; 1.
DR TIGRFSMS; TIGR01126; pdl.dom; 2.
DR PROSITE; PS00194; THIOREDOXIN; 2.
KW Protease; Redox-active center.
SQ SQUINCE 505 AA; 56678 MW; 3A7CD1C35981C4B3 CRC64;

Query Match 4.1%; Score 168.5; DB 11; Length 505;
Best Local Similarity 22.1%; Pred. No. 0.0013;
Matches 64; Conservative 55; Mismatches 125; Indels 45; Gaps 10;

OY 376 VELTEETFNATVNASDS---IVLFYAGQAVSMAPLOSYIDVAVKLTGSTMLLRINC 431
DB 28 LEITDENFESRVSDDTGSAGLMLVEFFAPWCCHKRLAPEYEAATRLKG--IYPLAKVDC 85
OY 432 ADMSDVCITKQNVTEEPPIKMYKKGENPVSYAGMLGTDLKLFIOLNRISTPVNITSIOEA 491
DB 86 TANTNTCNKYGVSQYPLTKIFRDGEAGAYDGPRTADGIVS--HLKKQAPASVPLRTEE 143
OY 492 EEYLSGELYKDLILYSSVYLGLFSPMTAKEDSEAGNYLKG---YVITGIYSEEDVL 548
DB 144 E-----FKKFSIDKQASVVGFRDLFSDGSHSEFLKAASNLKQNYFAHTNIES----- 191
OY 549 LLSKTYAASLPALLAR--HTEKIESIPLASTHAQ---DIVOITDALLEMPETIVE 602
DB 192 -LVKEYDNDGEGITIFRPLHLANKFEDKTVAYTEKKMTSGIKKFIQDSIFGLCPHMTED 250
OY 603 N-----LPSYFRLOKPLILFSDGTVPQYKKAILLVLYKOKYDS 642
DB 251 NKDLIOGKDLITAYVDV-----YKNAKGSNMYNRNRMVMAKFLDA 293

RESULT 11
O91281 PRELIMINARY: PRT; 505 AA.
ID O91281;
AC O91281;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

```

```

DE ERP57 protein.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RA Chung J.Y., Hwang S.O., Lee G.M.;
RC TISSUE=OVARY;
RA "The nucleotide sequence encoding the chinese hamster ERP57 protein.";
RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY057063; AAL18160.1; -.
DR InterPro; IPR000063; ThioRed.
DR Pfam; PF00085; thioRed; 2.
DR TIGRFSMS; TIGR01130; ER_PDI.fam; 1.
DR TIGRFSMS; TIGR01126; pdl.dom; 2.
DR PROSITE; PS00194; THIOREDOXIN; UNKNOWN.2.
SQ SEQUENCE 505 AA; 56796 MW; 5F2E38C30794DB76 CRC64;

Query Match 4.0%; Score 166.5; DB 11; Length 505;
Best Local Similarity 21.7%; Pred. No. 0.0017;
Matches 62; Conservative 57; Mismatches 128; Indels 39; Gaps 9;

OY 376 VELTEETFNATVNASDS---IVLFYAGQAVSMAPLOSYIDVAVKLTGSTMLLRINC 431
DB 28 LEITDENFESRVSDDTGSAGLMLVEFFAPWCCHKRLAPEYEAATRLKG--IYPLAKVDC 85
OY 432 ADMSDVCITKQNVTEEPPIKMYKKGENPVSYAGMLGTDLKLFIOLNRISTPVNITSIOEA 491
DB 86 TANTNTCNKYGVSQYPLTKIFRDGEAGAYDGPRTADGIVS--HLKKQAPASVPLRSEE 143
OY 492 EEYLSGELYKDLILYSSVYLGLFSPMTAKEDSEAGNYLKG---YVITGIYSEEDVL 551
DB 144 E-----FKKFSIDKQASVVGFRDLFSDGSHSEFLKAASNLND---NYRFAHTNIESLV 193
OY 552 TKTYAASLPALLAR--HTEKIESIPLASTHAQ---DIVOITDALLEMPETIVEN-- 603
DB 194 KEYDDNDEGEGITIFRPSHLANKFEDKTVVYEQMTSGIKRFOESIFGLCPHMTDNKD 253
OY 604 -----LPSYFRLOKPLILFSDGTVPQYKKAILLVLYKOKYDS 642
DB 254 LIOSKDLITAYVDV-----YKNAKGSNMYNRNRMVMAKFLDA 293

RESULT 12
O8R6N2 PRELIMINARY: PRT; 577 AA.
ID O8R6N2;
AC O8R6N2;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE ABC-type multidrug/protein/lipid transport system, ATPase
DE component.
GN MDIb12 OR TTE2769.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=MB4T/JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013214; AAM25873.1; -.
KW Complete proteome.
SQ SEQUENCE 577 AA; 66097 MW; BA6642DF1FBE7545 CRC64;

Query Match 3.9%; Score 162; DB 16; Length 577;

```

Best Local Similarity 19.6%; Pred. No. 0.0045;
Matches 142; Conservative 111; Mismatches 224; Indels 248; Gaps 33;

```

QY 82 MKAYLFKGNILLREPTDLEFVNVAIVAHVLFALLFSEVKYTTNEDLONINMALKGRKN 141
DB 1 MKEFLFK---YKKEFLVAVFLVAVSLINIVAFATYKRLIDVATRTDLVKFSYV--KFA 55
QY 142 IIFSIVRAIGIDENHRAVEAGFVYGT--YQFVLTEI---ALLSISEGEVYAHLYPF 196
DB 56 VMFTILEAV-----VGLNRTTRHRYMKKTLIYLAKDLFRALIKKDVH----- 99
QY 197 HCKVLVDLTQOCRTLMQPLTTLNHLFIKTKMAPLLEVAEDPOQV----- 244
DB 100 -----FNEVNTGKIYISINDYKIVYEDDYFNNFRL 131
QY 245 -STVHLQGLPLVFISSQATYEADRTAEWAMRLGAGVLLLRSLSEVNIQ--DA 301
DB 132 GSAGFVAAVLSLFTLSKTYT-----YMLIMALLSVIIPRIFD 171
QY 302 NVFEKRAEGVVEFLVLDVLIIS-----HYENNHIIEIOEDNDMEGPDIDVQ 354
DB 172 KIARKRNYSSELFTIESKDTLGLGVKSFGEIDKHEKFSKVED----- 220
QY 335 DDEVAETVFRDRKKLPLELTVELTEETFNATVMSDSIVLFYAGQAVSMAFLQSYDV 414
DB 221 -----VEDKKIKYSVLLN---TSDTMSSELSFIFLSVF-----AVGLYF----- 257
QY 415 AVKLKGTSMILTRINCADMSDVCTKQNTPEPIIKMYKKGKGNPVSAGMLGKTKLKI 474
DB 258 --TKGEMT-LGTMIACV-----QLTNNIIMPIYSMGON-----LNRTL 293
QY 475 QLNRIISYVNITSIOEAE---YLSGELYKDLILYSVSVGLFSPYTKTAKED-----F 526
DB 294 SLKTSIQKIN-EVLDKKEKENDYIPVKSNDSEIKRNS---FSTGTAKLDNINFTI 348
QY 527 SEAGNYLKVYITGTYSEEDVLLSTKYA--ASLPALLARHT---ECKIESIPLASTH 580
DB 349 KKGKRY-----ALVGTSGAGKSTILKLLQYENYEGETIKLDGIELRID 393
QY 581 AODIYQITDALLEMFPEITVENLPSYPRLOKPLILPSDGVNPOYKKAITLYKQKIL 640
DB 394 KKDLEKIIT-----LLHQN-----VFIFDGTV-----KDNITLLEDRYT 427
QY 641 DSFTPCMLNKTPYRGILRAYFDP-LPPLPLVLVNLHSGOYAFPSDQ-----AI 693
DB 428 DE-----EYVRAKTINGLGRPLEKLPDEGILSDVGECKLISGGERRIARSI 476
QY 694 IEEENLVMLKLEAGLENNH-----IT--ILPAQEMKPPPAYDFLSMT-DAA 737
DB 477 ITNASIILADEXTALDNETAYMIKTIIDMTIFAIVVTHRLMSSELKRYDELIYLRQGR 536
QY 738 TSORG 742
DB 537 IVEKG 541

```

RESULT 13

```

Q91815 PRELIMINARY: PRT: 369 AA.
AC Q91815;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Protein disulfide isomerase ER-60 (EC 5.3.4.1) (Fragment).
GN ERp60.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.

```

RA Fu R., Jiang M., Brenner S.;
RT "Structure of the ubiquitous mitochondrial creatine kinase gene and
its linkage to the protein disulfide isomerase ERp60 gene are
RT conserved among teleost fish Takifugu rubripes, mouse and human.";
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.

```

DR EMBL; AF274502; AAF78087.1; -.
DR HSSP; P07237; 1MER.
DR InterPro; IPR000063; ThioRed.
DR Pfam; PF00085; thioRed.
DR PRINTS; PR00421; THIOREDOXIN.
DR PROSITE; PS00194; THIOREDOXIN; 1.
KW Isomerase; Redox-active center.
FT VARIANT 116 116
FT NON_TER 369 369
FT SEQUENCE 369 AA; 41237 MW; 1438527CCAD01E55 CRC64;

```

Query Match 3.8%; Score 158; DB 13; Length 369;
Best Local Similarity 22.0%; Pred. No. 0.0043;
Matches 61; Conservative 60; Mismatches 130; Indels 26; Gaps 9;

```

QY 376 VELTEETFNATVMSDSIVL-FYAGQAVSMAFLQSYIDVAVKLKGTSMILTRINCAD 434
DB 21 LEFTDNDESKIGDHEITLVEFFAPWCGRKRLAPYERKAATALRG--VYPLAKVCTSN 78
QY 435 SDVCTKQNTPEPIIKMYKKGKGNPVSAGMLGKTKLKIQLNRIISYVNITSIOEAEY 494
DB 79 SNICKSYQVSGYPTLKVFDGEGSAYDGPRTSDGIVTFKKQYGPASVALAGEELQKF 138
QY 495 LSGELYKDLILYSVSVGLFSPYTKTAKEDSEAGNYLKG-YVITGTYSEEDVLLSTK 553
DB 139 ISEK-----DSSVGVFFADKSTAQVEFLKASALRDYXRA--HTNVEVLLKSON 187
QY 554 YVASLPALLARHTTEKIESIPLASTH---AODIYQITDALLEMFPEITVENLPSYR 609
DB 188 VQPEGIVLEPPPLTKKFFEDSSVKYSEKRYTSKIKRFIDQNLGRCPHMTEKDKQL-- 245
QY 610 LQKPLILPSDGTV--NPQ---YKKAITLYKQKYL 641
DB 246 MGKDLVAVYDVDERNPKGSNMWRMRVAKVAKTFED 282

```

RESULT 14

```

Q9GPH2 PRELIMINARY: PRT: 494 AA.
AC Q9GPH2;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Protein disulfide isomerase.
GN PDI.
OS Bombyx mori (Silk moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Bombycoidea; Bombycidae; Bombyx.
OX NCBI_TaxID=7091;
RN [1]
RP SEQUENCE FROM N.A.
RA Goo T.W., Yun E.Y., Hwang J.S., Kang S.W., Lee K.S., Lee J.S.,
RA Kwon O.-Y.;
RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF325211; AAG45936.1; -.
DR HSSP; P07237; 1MER.
DR InterPro; IPR00086; ER_target.
DR InterPro; IPR000063; ThioRed.
DR Pfam; PF00085; thioRed; 2.
DR PRINTS; PR00421; THIOREDOXIN.
DR TIGRfams; TIGR01130; ER_PDI_fam; 1.
DR TIGRfams; TIGR01126; pdi_dom; 2.
DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
DR PROSITE; PS00194; THIOREDOXIN; 2.
KW Isomerase.
RP SEQUENCE 494 AA; 55588 MW; 084609DD56DC88DD CRC64;

```

Query Match 3.78; Score 154; DB 5; Length 494;
 Best Local Similarity 22.08; Pred. No. 0.013;
 Matches 64; Conservative 54; Mismatches 125; Indels 48; Gaps 11;

QY 369 KLPELV-ELTEETFNATVWASDIYL-FYAGQAVSMALQSYIDVAVKL-KGTSMML 425
 DB 19 EVPEENVLVLSKNETFVISTTEYLVEFYAPMGCHCKSLAPEYAKAAKLAEESEPIK 78
 QY 426 LTRINCADWSDVCTKQNTKEPFIKMYKKGPNVSYAGMLCTKDLKFIQILNRISYPVNI 485
 DB 79 LAKVDATQEODLAESYGRGPLTKLFFRNG-SPIDYSGRQADDIISMLKKKTPPAVEY 137
 QY 486 TSIOAEELYSGLKYLKDLILYSSVVLGSPYTKTAKEDSEAGNLYKGYITIGYSEE 545
 DB 138 TSAQA-----KELIDANTVIVGFSDQSTRAKFTLSTAOVVDQVF-AIVSDE 187
 QY 546 DVLLSTKYAASLPALLARHTGKIESIPLASTHADIVQIITDALLEMPETVENLP 605
 DB 188 KVI-----KELEADEDVVLKMFEEKRYKYEDE-----ETTEDLNL 224

QY 606 SYFRLOK-PLLIFFSDGTVPD-----QYKAILTLVK---QRYLDSFTP 645
 DB 225 AMVAVQSMPTVERSHETASKIFGKIKYHLLIFLSKNGDFEKYLEDLKP 275

RESULT 15
 Q9C128
 ID Q9C128 PRELIMINARY; PRT; 517 AA.
 AC Q9C128;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Protein disulphide isomerase.
 GN PDI.
 OS Pichia pastoris (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Pichia.
 OX NCBI_TaxID=4922;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21139750; PubMed=11243858;
 RA Wiersma A.A., Vad R., Kristensen T., Oyen T.B.;
 RT "Characterization of a gene encoding a Pichia pastoris protein
 RT disulfide isomerase."
 RL Biochem. Biophys. Res. Commun. 281:1176-1182(2001).
 DR EMBL: AJ302014; CAC33587.1; -.
 DR HSSP: P07237; IMEX.
 DR InterPro: IPR000886; ER target.
 DR InterPro: IPR000063; Thioled.
 DR Pfam: PF000085; thioled; 2.
 DR PRINTS: PR00421; THIOREDOXIN.
 DR TIGRFAMS: TIGR01130; ER_PDI_fam; 1.
 DR TIGRFAMS: TIGR01126; pdi_dom; 2.
 DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
 DR PROSITE: PS00194; THIOREDOXIN; 2.
 KW isomerase; Redox-active center.
 SQ SEQUENCE 517 AA: 57788 MW; 4C085C534B65583 CRC64;

Query Match 3.6%; Score 150.5; DB 3; Length 517;
 Best Local Similarity 22.5%; Pred. No. 0.025;

Matches 79; Conservative 64; Mismatches 161; Indels 47; Gaps 15;

QY 376 VELTEETFNATVWASDIYL-FYAGQAVSMALQSYIDVAVKLKGTSTMLLTRINCADW 434
 DB 36 VKLTEAFEFESFITSNPHVLAFFAPMGCHCKILGPELSAAELIKDNEQVIAQIDCTEE 95
 QY 435 SDVCTKQNTKEPFIKMYK-KGENPVSYAGMLCTKDLKFIQILNRISYPVNIISIQEAE 493
 DB 96 KELCOGEIYGYPPLKVFHGEVEPSPDYQOQSOSIVSYMLKQSLPPVSEINATKDLD 155
 QY 494 YLSGELYKDLILYSSVVLGSPYTKTAKEDSEAGNLYKGYITIGYSEE--DVLLLS 551
 DB 156 TIA-EAKEPVI---VOVL-----PEDASNLESNTTFYGVAGTLREKFTFVSTKS 200

QY 552 TKYAA-----SLPALLARHTGKIESI-----PLASTHADIVQIITDALLEMPETIVE 602
 DB 201 TDYAKKRTSDSTPAYILVR--PGEEPSVYSGEELDETHLVHWDIESKPL---FGDIDGS 255
 QY 603 NLPSYFRLOKPLLIFFSDGTVPDQYKAILTLVK-----QRYLDSFTPCLNLRKNTFVGR 657
 DB 256 TFKSYAANIPLAYFYE---NEQRAAADAIIKPKFAKQORGKINF---VGLDAVKEGK 308
 QY 658 GILRAYDPPLPPLVLVNLHSGQGVFAFPDQALIEENLVMLKLEAG 708
 DB 309 HAKNLNME-EKLPFYIHDLVS-NKKEGVPODELTKNVTLEIKFTIAG 357

Search completed: March 13, 2003, 17:14:05
 Job time : 41 secs

| | | | |
|----|------|--|------|
| Db | 2401 | AATATCACTTGATATCGTGTCTTAAATTTCTTCACATATGGTTGGCCGATACAAAC | 2460 |
| OY | 4146 | TGATTTTATTAACGTAAATTTAAGAAATCTAACAGCTAAACTCAGTAATGCAATTAAT | 4205 |
| Db | 2461 | TGATTTTATTAACCTAAATTTAAGAAATCTAACAGCTAAACTCAGTAATGCAATTAAT | 2520 |
| OY | 4206 | TCCTTATTAACATAGACCCGTTGCTACTCTCAGACCCCTCCCTCAATTTTTCCTGTA | 4265 |
| Db | 2521 | TCCTTATTAACATAGACCCGTTGCTACTCTCAGACCCCTCCCTCAATTTTTCCTGTA | 2580 |
| OY | 4266 | GCATGTGATGCGCTGATTTAACTCATTTTCATTTGCTTTATTTCTTAATATGGGAACAATG | 4325 |
| Db | 2581 | GCATGTGATGCGCTGATTTAACTCATTTTCATTTGCTTTATTTCTTAATATGGGAACAATG | 2640 |
| OY | 4326 | AGAGTGAACCTCAATATATAGGTGTAGTAATTAACAATCATATAGCCTAATTTTGAANA | 4385 |
| Db | 2641 | AGAGTGAACCTCAATATATAGGTGTAGTAATTAACAATCATATAGCCTAATTTTGAANA | 2700 |
| OY | 4386 | TGCTTAATTAAGTACGACGACATAGAACAATGCAATTTGCTTAAGTCAATTTGACCTTTGTCAG | 4445 |
| Db | 2701 | TGCTTAATTAAGTACGACGACATAGAACAATGCAATTTGCTTAAGTCAATTTGACCTTTGTCAG | 2760 |
| OY | 4446 | CAATTTTGACAGTCATTTATGTGTTGTCATTAATTTTAAATTAAGGTGTCGGGTTTCAGAT | 4505 |
| Db | 2761 | CAATTTTGACAGTCATTTATGTGTTGTCATTAATTTTAAATTAAGGTGTCGGGTTTCAGAT | 2820 |

| | |
|------------|---|
| RESULT | 3 |
| AKO97748 | |
| LOCUS | 2747 bp mRNA linear PRI 15-JUL-2002 |
| DEFINITION | Homo sapiens CDNA FLJ40429 fls, clone TEST12039177. |
| ACCESSION | AKO97748 |
| VERSION | AKO97748.1 GI:21757613 |
| KEYWORDS | Oligo capping; fls (full insert sequence). |
| SOURCE | Homo sapiens testis CDNA to mRNA, clone_11b:TEST12 |

CONSUMISM
nolus sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
1
Oshima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K.,

Arita, M., Mutsashino, K., Yuuchi, H., Hara, H., Sugiyama, T., Irie, R.,
 Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,
 Kawai-Hito, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,
 Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,
 Nagatsuna, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B.,
 Suzuki, T., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and
 Isogai, T.

| | |
|------------------|--|
| JOURNAL TITLE | NEDO human cDNA sequencing project Unpublished 2 (bases 1 to 2747) |
| REFERENCE | Isogai, T., and Yamamoto, J. |
| AUTHORS | Direct Submission |
| JOURNAL | Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team): 2-6-7 |

COMMENT

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TEST12039177"
/tissue_type="testis"
/clone_lib="TEST12"
/note="cloning vector: PME18SFL3"

```

| BASE COUNT | 899 a | 510 c | 575 g | 763 t |
|------------|-------|-------|-------|-------|
| ORIGIN | | | | |

| | | | | |
|-----------------------|--------------|---------------|---------------|--------------|
| Query Match | 53.3% | Score 2410.2; | DB 9; | Length 2747; |
| Best Local Similarity | 94.2%; | Pred. No. 0; | | |
| Matches 2588; | Conservative | 0; | Mismatches 3; | Indels 156; |
| | | | | Gaps 2 |

| | | | |
|----|------|---|------|
| Oy | 139 | CTTGCCCGCCGCGCCCTCGGCAAGCGCCCGAGTCCCGCCCGCAGGTCCTCCGGGCAGATTAAC | 198 |
| Db | 1 | GTTGCGCCGCGCGCGCCCTCGGGAACGGCCAGGTCCCGCCCGCGAGGTCCCGGGGAGATTAAC | 60 |
| Oy | 199 | ATAGATCATCGTGTGAAGAACTCTCTTGGAAGTTGTTCGAAGAAATTTGGAAGTGTGCAAAAT | 258 |
| Db | 61 | ATAGATCATCTGCTGAAAACCTCTTGGAAGTTGTTCGAAGAAATTTGGAAGTGTGCAAAAT | 120 |
| Oy | 259 | AGAAATTAAGAATTTAACAGCAGATACAGAGAGACAGCATGGAAGTGTGTCTTAGGAAC | 318 |
| Db | 121 | AGAAATTAAGAATTTAACAGCAGATACAGAGAGACAGCATGGAAGTGTGTCTTAGGAAC | 180 |
| Oy | 319 | AGAACACGACGAGTAAAAAACAGACAAATTCCTCGATACAACTGACGCTGATATATGT | 378 |
| Db | 181 | AGAACACGACGAGTAAAAAACAGACAAATTCCTCGATACAACTGACGCTGATATATGT | 240 |
| Oy | 379 | TTTCCGGCTCAATGTCTTTAGATTTGGATTCCTTTGTCAATAATGTCAATTTTACA | 438 |
| Db | 241 | TTTCCGGCTCAATGTCTTTAGATTTGGATTCCTTTGTCAATAATGTCAATTTTACA | 300 |
| Oy | 439 | TGCCAACGTAACTCTTTACGAGACTGAGTCTCAGAAATATTTTACTACTTGCAC | 498 |
| Db | 301 | TGCCAACGTAACTCTTTACGAGACTGAGTCTCAGAAATATTTTACTACTTGCAC | 360 |
| Oy | 499 | CAGC-----TC | 504 |
| Db | 361 | CAGGAAGAGCTCTTATAGCTATTTTGTCAACCTGATTCCTCCCAACACATCTGATATTC | 420 |
| Oy | 505 | TTGAGAACGTGATGAGGCTGTTAGACTCTGCAGGACTATGGAATTTAGTTGCCAACG | 564 |
| Db | 421 | TTGAGAACGTGAAAGAGGCTGTTAGACTCTGCAGGACTATGGAATTTAGTTGCCAACG | 480 |
| Oy | 565 | TTAATGTGTCAAAGAAGAAATATCAAGATACGTGTGAAAAAGAAAGATTTGATGAAG | 624 |
| Db | 481 | TTAATGTGTCAAAGAAGAAATATCAAGATACGTGTGAAAAAGAAAGATTTGATGAAG | 540 |
| Oy | 625 | CATATTTATTCAGGGCACATATTTGCTCAGAGAAATTCCTACTGCACCTGTGTTTGATG | 684 |
| Db | 541 | CATATTTATTCAGGGCACATATTTGCTCAGAGAAATTCCTACTGCACCTGTGTTTGATG | 600 |
| Oy | 685 | TGAATGCCATTTGTGCCCATGTCTCTCTTCTCTTTTAACTGAAGTGAATATATTA | 744 |
| Db | 601 | TGAATGCCATTTGTGCCCATGTCTCTCTCTTTTAACTGAAGTGAATATATTA | 660 |
| Oy | 745 | CCAACCTGGAAGACCTTCAGACATATGAAGAAATCTCTGAAGAGGAAGCAAAATATTATAT | 804 |
| Db | 661 | CCAACCTGGAAGACCTTCAGACATATGAAGAAATCTCTGAAGAGGAAGCAAAATATTATAT | 720 |
| Oy | 805 | TCTCATTTGTAAGAGCCATTGGGAATACAGAGACACAGAGCAGTCATGGAAGCCGTTTTG | 864 |
| Db | 721 | TCTCATTTGTAAGAGCCATTGGGAATACCAAGACACAGAGCAGTCATGGAAGCCGTTTTG | 780 |
| Oy | 865 | TGTTTGGAGCTACTACTACCAATTTGTCTTAACACAGAAATGGCCTTTTGAAGATATTG | 924 |
| Db | 781 | TGTTTGGAGCTACTACTACCAATTTGTCTTAACACAGAAATGGCCTTTTGAAGATATTG | 840 |
| Oy | 925 | GCTCTGAGAGTGTGGAAATATGCACATCTCTACTTTTTCATTTGTAACATGATGCTTGACT | 984 |
| Db | 841 | GCTCTGAGAGTGTGGAAATATGCATATCTCTACTTTTTCATTTGTAACATGATGCTTGACT | 900 |
| Oy | 985 | TGACCCAGCAATGTAGAAAGAACTAATGGAACAGCCATTGACTACAGCAATTCAC | 1044 |
| Db | 901 | TGACCCAGCAATGTAGAAAGAACTAATGGAACAGCCATTGACTACAGCAATTCAC | 960 |
| Oy | 1045 | TGTTTATTAAGACATGAAAGACCTCTGTTGACGTGAAGTGTCTGGAAGTCTCTCAACAG | 1104 |

THIS PAGE BLANK (USPTO)

DB 961 TGTATTAAGCAATGAAAGCACTGTGTAGTGAAGTGTCTGAAGATCTCTCAACAG 1020
OY 1105 TTTCACTGCTCCATCTCCAACTGGGCTTACCACTGGTTTTTATTTAGTCAACAGCTA 1164
DB 1021 TTTCACTGCTCCATCTCCAACTGGGCTTACCACTGGTTTTTATTTAGTCAACAGCTA 1080
OY 1165 CTATGAGCTGTATAGAAAGTCAAGTGGTTCCTTGGCTCTTGGGAAAGCAG 1224
DB 1081 CTATGAGCTGTATAGAAAGTCAAGTGGTTCCTTGGCTCTTGGGAAAGCAG 1140
OY 1225 GAGTCTACTCTTGTTAAGGAGCTTTTGGAGTCAACATCTCTCAAGATGCTAATGG 1284
DB 1141 GAGTCTACTCTTGTTAAGGAGCTTTTGGAGTCAACATCTCTCAAGATGCTAATGG 1200
OY 1285 TCTTCAAAAGAGCAAGAGGAGTTCAGTGAATTTTGTATTACATGATGTGATT 1344
DB 1201 TCTTCAAAAGAGCAAGAGGAGTTCAGTGAATTTTGTATTACATGATGTGATT 1260
OY 1345 TAAATATATCTCATGTGGAAATATATGCAATGAGAAATTCAGAAAGATGAGACA 1404
DB 1261 TAAATATATCTCATGTGGAAATATATGCAATGAGAAATTCAGAAAGATGAGACA 1320
OY 1405 ATGCAATGGAAGGCTCCAGATATGATGTGAGATGATGAGATGAGAACTGTTTCA 1464
DB 1321 ATGCAATGGAAGGCTCCAGATATGATGTGAGATGATGAGATGAGAACTGTTTCA 1380
OY 1465 GAGATAGGAAGAGAAATTAACCTTTGGAACCTTACAGTGAACATTAACAGAAACATTTA 1524
DB 1381 GAGATAGGAAGAGAAATTAACCTTTGGAACCTTACAGTGAACATTAACAGAAACATTTA 1440
OY 1525 ATGCAACAGTATGGCTTCTGACAGCATAGTACTCTTATGCGTGGTGGCAGCAGAT 1584
DB 1441 ATGCAACAGTATGGCTTCTGACAGCATAGTACTCTTATGCGTGGTGGCAGCAGAT 1500
OY 1585 CCATGGCATTTTTCGAATCTATATTTGATGTGAGATGATGAGATGAGAACTGTTTCA 1631
DB 1501 CCATGGCATTTTTCGAATCTATATTTGATGTGAGATGATGAGATGAGAACTGTTTCA 1560
OY 1632 ----- 1631
DB 1561 ATCTTCTGCTCTTACAGCACATGCTGCGAGAGAGGAGAGAGAGAGAGAGAGAGAG 1620
OY 1632 -----GGCAGATCTACTATGCTTCTTACTAGATTA 1662
DB 1621 TTTGAAACGTTAGTCTTACTTCTTCTGCGGACATCTACTATGCTTCTTACTAGATTA 1680
OY 1663 ACTGTCAGATTTGCTGTATGTATGTATGTAAGCAAAATGTTACTGAATTTCTATCATTA 1722
DB 1681 ACTGTCAGATTTGCTGTATGTATGTATGTAAGCAAAATGTTACTGAATTTCTATCATTA 1740
OY 1723 AGATGTACAGAAAGGCGAGAACCCAGTATCTTATGCTGGAATGTTAGGAACCGAAGATC 1782
DB 1741 AGATGTACAGAAAGGCGAGAACCCAGTATCTTATGCTGGAATGTTAGGAACCGAAGATC 1800
OY 1783 TCCATAATTTATCCAGTCAACAGATTTTCAATCCAGTGAATTAACATGATCCAG 1842
DB 1801 TCCATAATTTATCCAGTCAACAGATTTTCAATCCAGTGAATTAACATGATCCAG 1860
OY 1843 AAGCAGAAATATTTAAGTGGGGAATTAATTAAGACCTCATCTTGTATTTAGTGTCT 1902
DB 1861 AAGCAGAAATATTTAAGTGGGGAATTAATTAAGACCTCATCTTGTATTTAGTGTCT 1920
OY 1903 CAGTATTTGGGACTATTTAGTCCACCATGAAACAGCAAAAGAGATTTTATGAGAGCAG 1962
DB 1921 CAGTATTTGGGACTATTTAGTCCACCATGAAACAGCAAAAGAGATTTTATGAGAGCAG 1980
OY 1963 GAAATCTACTAAAGATATGTTATATCACTGGAATTTATCTGAAGATGTTTGTCTAC 2022
DB 1981 GAAATCTACTAAAGATATGTTATATCACTGGAATTTATCTGAAGATGTTTGTCTAC 2040
OY 2023 TGTCAACCAAAATATGCTGCAAGCTTCCAGACCTGCTGCTTCCAGACAGACAGAGAGCA 2082
DB 2041 TGTCAACCAAAATATGCTGCAAGCTTCCAGACCTGCTGCTTCCAGACAGACAGAGAGCA 2100

OY 2083 AATAGAGAGCATCCACCTAGCTAGACACATGCAACAGATAGTCAAAATATTAACAG 2142
DB 2101 AATAGAGAGCATCCACCTAGCTAGACACATGCAACAGATAGTCAAAATATTAACAG 2160
OY 2143 ATGCACTACTGGAATGTTCCGGAATCACTGTGGAATATCTCCACTATTTTCAGAC 2202
DB 2161 ATGCACTACTGGAATGTTCCGGAATCACTGTGGAATATCTCCACTATTTTCAGAC 2220
OY 2203 TTCAGAAACCATATTTATTTGTTGTCAGTGAAGCAGCTGAATCTCAATATTAAGAG 2262
DB 2221 TTCAGAAACCATATTTATTTGTTGTCAGTGAAGCAGCTGAATCTCAATATTAAGAG 2280
OY 2263 CAATATTGACACTGTGAAGCAGAAATCTGATTCATTTACTCATCTGCTGTTAAATC 2322
DB 2281 CAATATTGACACTGTGAAGCAGAAATCTGATTCATTTACTCATCTGCTGTTAAATC 2340
OY 2323 TAAAGATATCTCCAGTGGGAGAGAACTTGGGCAATTTTGAATCTCTGCTGCTCC 2382
DB 2341 TAAAGATATCTCCAGTGGGAGAGAACTTGGGCAATTTTGAATCTCTGCTGCTCC 2400
OY 2383 TTTCTCTCTTGTGTTTGGTGAATCTGATTCAGTGGGCAATTTTGAATCTCTGCTGCTCC 2442
DB 2401 TTTCTCTCTTGTGTTTGGTGAATCTGATTCAGTGGGCAATTTTGAATCTCTGCTGCTCC 2460
OY 2443 ACCAGCTATATTTGAAGAAACCTTGTATTTGGCTGAAGAAATTAAGAGCAGACTAG 2502
DB 2461 ACCAGCTATATTTGAAGAAACCTTGTATTTGGCTGAAGAAATTAAGAGCAGACTAG 2520
OY 2503 AAAATCATATCAATTTTACCTGCTCAAGATGGAACCTCTCTCCAGCTTATGATT 2562
DB 2521 AAAATCATATCAATTTTACCTGCTCAAGATGGAACCTCTCTCCAGCTTATGATT 2580
OY 2563 TTTCAATATGATGATGATGCGGCAACATCTCAAGCTGACATGAGAAAGTCCCAAGTGA 2622
DB 2581 TTTCAATATGATGATGATGCGGCAACATCTCAAGCTGACATGAGAAAGTCCCAAGTGA 2640
OY 2623 TGAAGAAACAGATGTCCAGAGAGATGATGAAGAACATGAATTAATTCGCACTCA 2682
DB 2641 TGAAGAAACAGATGTCCAGAGAGATGATGAAGAACATGAATTAATTCGCACTCA 2700
OY 2683 GAAAGAAACGATTTGAAGCTGTGAATTAAGCACTTGAATTAAGAGT 2729
DB 2701 GAAAGAAACGATTTGAAGCTGTGAATTAAGCACTTGAATTAAGAGT 2747

RESULT 4
AC025659/c 161100 bp DNA linear HTG 25-JUN-2000
LOCUS Homo sapiens clone Rpl1-263N17, WORKING DRAFT SEQUENCE, 15
DEFINITION Unordered pieces.
ACCESSION AC025659
VERSION AC025659.3 GI:8705128
KEYWORDS HTG; HTGS; PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 161100)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens, clone Rpl1-263N17
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 161100)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abrahams, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campoliano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Colliore, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J., S., Dodge, S., Domino, M., Doyle, M., Ferrelita, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J., C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Labroque, R., Lamazares, R., Landers, T., Lehoczy, J.,

THIS PAGE BLANK (USPTO)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

THIS PAGE BLANK (USPTO)

QY 2684 AAAGAACCCGATTGAACCTCTGAGAAATGAAGCAATTTGAATAGAGTAATTTGGTTTAAAGA 2743
 DB 1038 AAAGAACCCGATTGAACCTCTGAGAAATGAAGCAATTTGAATAGAGTAATTTGGTTTAAAGA 1097
 QY 2744 ACCAGAAAATCATTTAGACGCTGATAAAGAGTTAGAGATGCTCAAAAGTGAACCTAAATTTTA 2803
 DB 1098 AGCAGAAAATCATTTAGACGCTGATAAAGAGTTAGAGATGCTCAAAAGTGAACCTAAATTTTA 1157
 QY 2804 TAGGCTGTGGTTTCCAAAATTTTGGCATGATAGATTAATTTATTTTCCCTTAAAGAA 2863
 DB 1158 TAGGCTGTGGTTTCCAAAATTTTGGCATGATAGATTAATTTATTTTCCCTTAAAGAA 1217
 QY 2864 TAATATTAATCATTTCAAGCTTGCAGACTAGTGCATCAATAGAGTAATTAATATAAGT 2923
 DB 1218 TAATATTAATCATTTCAAGCTTGCAGACTAGTGCATCAATAGAGTAATTAATATAAGT 1277
 QY 2924 CACATATTTTAAATTTTCTAGTAACATTAACCAAAAGTAAAAGTGAAGGAGGC 2983
 DB 1278 CACATATTTTAAATTTTCTAGTAACATTAACCAAAAGTAAAAGTGAAGGAGGC 1337
 QY 2984 AAAATTAATTTTGAATTTTCAACCCAGTACTATACCCAAATAGCGAAATATAGAA 3043
 DB 1338 AAAATTAATTTTGAATTTTCAACCCAGTACTATACCCAAATAGCGAAATATAGAA 1397
 QY 3044 TTATTAATGAGATATTTTACATCTTTTGTACCAAGTCTTCTAAATGACGATACATTT 3103
 DB 1398 TTATTAATGAGATATTTTACATCTTTTGTACCAAGTCTTCTAAATGACGATACATTT 1457
 QY 3104 TTATTAATGAGATATTTTACATCTTTTGTACCAAGTCTTCTAAATGACGATACATTT 3163
 DB 1458 TTATTAATGAGATATTTTACATCTTTTGTACCAAGTCTTCTAAATGACGATACATTT 1517
 QY 3164 TGGCTGTGACTACTGATTTGGAGCTGAGTACTAGCAAAAACCTAGCATTAATTAACCT 3223
 DB 1518 TGGCTGTGACTACTGATTTGGAGCTGAGTACTAGCAAAAACCTAGCATTAATTAACCT 1577
 QY 3224 AGTTCTAGCCATGATTTCTATTTGGATTAATTAATTAATTAATTAATTAATTAATTAAT 3283
 DB 1578 AGTTCTAGCCATGATTTCTATTTGGATTAATTAATTAATTAATTAATTAATTAATTAAT 1637
 QY 3284 GTGCAATTCATGAGCTGACAGTATATTTGTTTATTTGGAGTATGATATTAATTAATTAAT 3343
 DB 1638 GTGCAATTCATGAGCTGACAGTATATTTGTTTATTTGGAGTATGATATTAATTAATTAAT 1697
 QY 3344 GTTGTCAACCTCAGGCGATATTTAGCAATTTGCGGAGACATTTTGTGATGCTATGACTA 3403
 DB 1698 GTTGTCAACCTCAGGCGATATTTAGCAATTTGCGGAGACATTTTGTGATGCTATGACTA 1757
 QY 3404 GGGCAGTTATTTGACATTTAGTGTAGAGCCATGATCTGCTAAATTAATTAATTAATTAATTAAT 3463
 DB 1758 GGGCAGTTATTTGACATTTAGTGTAGAGCCATGATCTGCTAAATTAATTAATTAATTAATTAAT 1817
 QY 3464 ACAGCGCCCAACAAGAAATTTATCTGCGGAAATGTTAGTCTGCGCAAGGCTGAGTA 3523
 DB 1818 ACAGCGCCCAACAAGAAATTTATCTGCGGAAATGTTAGTCTGCGCAAGGCTGAGTA 1877
 QY 3524 ACCTTGTGTTAAAGTAACCTGTGGCAGACTAGTGTTCAGAAATTTTCCAGTATTTTCCGCTCA 3583
 DB 1878 ACCTTGTGTTAAAGTAACCTGTGGCAGACTAGTGTTCAGAAATTTTCCAGTATTTTCCGCTCA 1937
 QY 3584 CGTATCATGTTTGAATAATTTTGGCTATTTAAAGATATGATTTAGATGGTCTTATCTCTGA 3643
 DB 1938 CGTATCATGTTTGAATAATTTTGGCTATTTAAAGATATGATTTAGATGGTCTTATCTCTGA 1997
 QY 3644 TTATTAATGATTAACATTTGATCTTTTCTAAATTTTCAAGATGATGGGATAACCCCT 3703
 DB 1998 TTATTAATGATTAACATTTGATCTTTTCTAAATTTTCAAGATGATGGGATAACCCCT 2057
 QY 3704 AGAAGAGGACTCAGATGATATTTATTTTAAAGTGAAGTCTTAAACCTCTCTTATTTTC 3763
 DB 2058 AGAAGAGGACTCAGATGATATTTATTTTAAAGTGAAGTCTTAAACCTCTCTTATTTTC 2117

QY 3764 TACAAGTTATATGGCTAAATTTTTCAGATTTGAACAGGATTCAGCATTTCTGCCATCTCTCTCA 3823
 DB 2118 TACAAGTTATATGGCTAAATTTTTCAGATTTGAACAGGATTCAGCATTTCTGCCATCTCTCTCA 2177
 QY 3824 TGGAAAGAGAGGCTCCCTCATCTGAAAGCGTCTCTGAAATCTACCCCTTGAAGCTTCAGAC 3883
 DB 2178 TGGAAAGAGAGGCTCCCTCATCTGAAAGCGTCTCTGAAATCTACCCCTTGAAGCTTCAGAC 2237
 QY 3884 AAATCAGTTGATCTCCCTGAGCCACAGCGCTCATCTGAGGGAGGAGGAAAGATTAAGCC 3943
 DB 2238 AAATCAGTTGATCTCCCTGAGCCACAGCGCTCATCTGAGGGAGGAGGAAAGATTAAGCC 2297
 QY 3944 AAAGAGTTAAATTTTCAATTTCCAAATCACTTACCTGATCTGTTTGTAGCAGTTG 4003
 DB 2298 AAAGAGTTAAATTTTCAATTTCCAAATCACTTACCTGATCTGTTTGTAGCAGTTG 2357
 QY 4004 TTTGTCTCATTTTGTCTGTCATTTTTCAGACATTTTTCAGACATATTTCTATTGTT 4063
 DB 2358 TTTGTCTCATTTTGTCTGTCATTTTTCAGACATTTTTCAGACATATTTCTATTGTT 2417
 QY 4064 GCTCTACTGATTTTCTTTTAAATATCTACTGATCTGTTTCTTTTAAATTTTCTCTCA 4123
 DB 2418 GCTCTACTGATTTTCTTTTAAATATCTACTGATCTGTTTCTTTTAAATTTTCTCTCA 2477
 QY 4124 CATATGTTTCCCTGATACAACTGATTTTATTAAGTAAATTTAAGGAATCTAACAGCTA 4183
 DB 2478 CATATGTTTCCCTGATACAACTGATTTTATTAAGTAAATTTAAGGAATCTAACAGCTA 2536
 QY 4184 AAATCAGTAAGTGCAATTTTCTTATTAACATAGACCCGTTGCTACTCTCAGCACCCCT 4243
 DB 2537 AAATCAGTAAGTGCAATTTTCTTATTAACATAGACCCGTTGCTACTCTCAGCACCCCT 2592
 QY 4244 CTCTCAATTTTCTCTGATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 4303
 DB 2593 CTCTCAATTTTCTCTGATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 2647
 QY 4304 TATTCTAATATGGAACAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4363
 DB 2648 TATTCTATCTGTTGGAGCCCTGGAGGACCTTAATTTGGGGTGGGTATATAAACATCTT 2707

RESULT 4
 ABV27799
 ID ABV27799 standard; cDNA; 2905 BP.
 AC ABV27799;
 XX
 AC ABV27799;
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 27790.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US05171.
 XX
 PR 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.
 PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX

THIS PAGE BLANK (USPTO)
